

Implementación de un Data Mart como una solución de inteligencia de negocios aplicando metodología ágil en base a la plataforma académica MOODLE de la Universidad Tecnológica de Panamá

Implementation of a Data Mart for the MOODLE virtual academic platform as a business intelligence solution applying agile methodology at the Technology University of Panama

Danitzel Jurado ¹, Ramfis Miguelena ^{2*}

¹ Licenciatura en Tecnología de Programación y Análisis de Sistema, Facultad de Ingeniería de Sistemas Computacionales – Universidad Tecnológica de Panamá; ² Departamento de Sistemas de Información, Control y Evaluación de Recursos Informáticos – Facultad de Ingeniería de Sistemas Computacionales – Universidad Tecnológica de Panamá

Resumen El presente proyecto consiste en implementar un Data Mart, tecnología que forma parte de la inteligencia de negocios, aplicando metodología ágil BEAM* (Business Event Analysis & Modeling) para el diseño de Data Mart. Los datos utilizados como fuentes de información se obtendrán de la base de datos de la plataforma virtual académica MOODLE de la Universidad Tecnológica de Panamá (UTP), con el propósito de optimizar el proceso de extracción y transformación de datos en información relevante y útil que sirvan de apoyo en la toma de decisiones estratégicas a los directivos de la Universidad Tecnológica de Panamá.

Palabras clave Metodología ágil, BEAM*, almacén de datos, Data Mart, inteligencia de negocios.

Abstract The present project consists of implementing a Data Mart, technology that is part of the Business Intelligence, applying agile BEAM* methodology (Business Event Analysis & Modeling) for the design of Data Mart, the data used as information sources will be obtained of the database of the academic virtual platform MOODLE of the Technological University of Panama (UTP), with the purpose of optimizing the process of extracting and transforming data into relevant and useful information that will serve as support in the making of strategic decisions the directors of the Universidad Tecnológica de Panamá.

Keywords Methodology agile, BEAM, data warehouse, Data Mart, business intelligence

* Corresponding author: ramfis.miguelena@utp.ac.pa

1. Introducción

El uso de la inteligencia de negocios es vital en las organizaciones, ya que nos encontramos en la era de la información o era digital, periodo que inició en 1975 y se ha caracterizado, por la recolección y transmisión casi instantánea de grandes cantidades de información y por el auge de las industrias basadas en la información.

En la actualidad la inteligencia de negocios juega un papel importante en la toma de decisiones. Forrester research indicó que solo un pequeño porcentaje de los datos de las organizaciones se convierten realmente en información útil al servicio de una mejor comprensión y toma de decisiones [1].

El propósito de este proyecto es implementar un Data Mart utilizando metodología ágil para explorar y explotar los datos

utilizando la tabla de eventos o *logs* que utiliza la plataforma académica MOODLE, con la finalidad de medir el comportamiento de los usuarios (estudiantes y docentes), para la toma de decisiones de los directivos de la UTP.

2. Aspectos Generales

2.1 Antecedentes

Los Entornos Virtuales de Aprendizaje (EVA) conocidos también como Learning Management System (LMS), son aplicaciones que sirven de apoyo en la formación y mejoramiento continuo de los docentes y estudiantes.

Un EVA es “Un conjunto de herramientas informáticas que posibilitan la interacción didáctica de manera que el alumno

pueda llevar a cabo las labores propias de la docencia como son conversar, leer documentos, realizar ejercicios, formular preguntas al docente, trabajar en equipo, entre otros” [2].

La UTP implementó la plataforma virtual Modal Object Oriented Dynamic Learning Environment (MOODLE) en el año 2001, como método alternativo de enseñanza para profesores y estudiantes, que por compromisos personales y laborales no han podido continuar con su formación universitaria o mantenerse actualizados. Esta plataforma estuvo bajo la administración del centro de investigación UTPVIRTUAL hasta el año 2010, específicamente el 10 de mayo, donde el centro de UTPVIRTUAL se convierte en el Centro de Investigación, Desarrollo e Innovación en Tecnologías de la Información y las Comunicaciones (CIDITIC). A partir del segundo semestre del año 2018, la administración de la plataforma virtual MOODLE está a cargo de la Dirección General de Innovación y Tecnología Educativa (DIGITED) [3].

MOODLE, es una plataforma virtual (son aplicaciones orientadas al internet que se utilizan para desarrollar cursos o módulos didácticos en la red internacional), de código libre, desarrollada en el lenguaje PHP (Hypertext Preprocessor-Procesador de hipertexto) basada en una determinada filosofía del aprendizaje, una forma de pensar que a menudo se denomina “pedagogía constructorista social” [4].

Características principales de la plataforma:

- Basada en la pedagogía constructivista social (colaboración, actividades, reflexión crítica, etc.).
- Fue registrado como licencia pública GPL (GNU), es decir es un *software* libre de uso.
- Tiene una interfaz moderna, fácil de utilizar y compatible.
- Apropia para el 100% de clases en línea.

2.2 Caracterización de la problemática

La plataforma académica virtual MOODLE denominada E-CAMPUS de la UTP, genera una gran cantidad de datos, producto de las interacciones realizadas por los usuarios de la plataforma. Estos datos requieren ser explotados y analizados con el propósito de obtener información útil para generar conocimientos que sirvan de apoyo en la toma de decisiones.

En la actualidad el personal del Centro de Investigación CIDITIC, realiza extracciones de datos para transformarlos en información y entregar resultados para solicitudes específicas que sirven de apoyo para la toma de decisiones, pero el tiempo que se requiere para realizar este proceso no es óptimo ni oportuno, al menos se requiere de una (1) semana para realizar la entrega de la información solicitada.

En la figura 1, se detalla el proceso que realiza el personal del Centro CIDITIC.

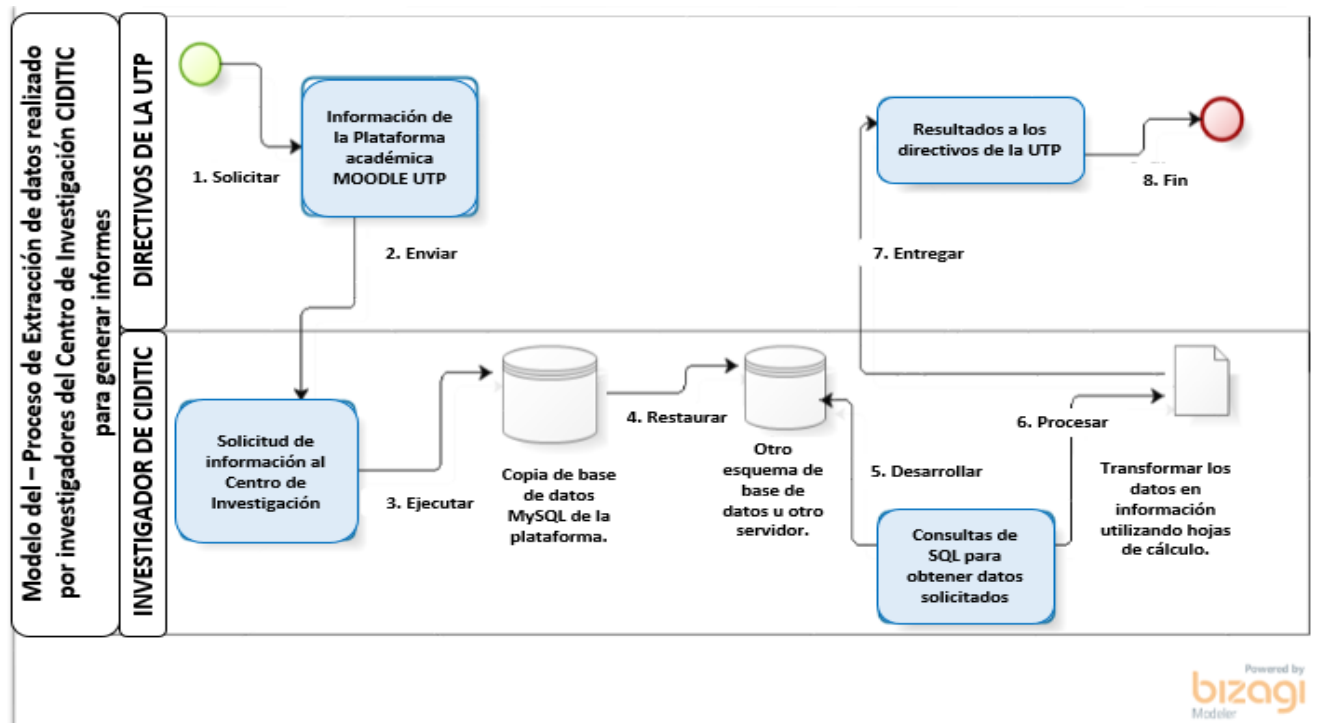


Figura 1. Presenta el modelo del proceso que realiza el personal del centro CIDITIC para extraer datos de la plataforma académica virtual MOODLE UTP y entregar resultados de solicitudes específicas.

Fuente: Elaboración propia.

2.3 Objetivos

2.3.1 Objetivo general

Implementar un Data Mart aplicando inteligencia de negocio a la plataforma académica MOODLE de la Universidad Tecnológica de Panamá, utilizando metodologías de desarrollo de *software* ágil BEAM*.

2.3.2 Objetivos específicos

- Aplicar los fundamentos de la inteligencia de negocios en el área académica para la toma de decisiones oportunas.
- Analizar la estructura de datos de la plataforma académica virtual de MOODLE para identificar la fuente de datos a utilizar y poder diseñar el modelo lógico del Data Mart.
- Diseñar el modelo lógico para crear el modelo físico del repositorio central del Data Mart.
- Seleccionar la herramienta de extracción, transformación y carga de los datos para automatizar el proceso.
- Construir el Data Mart para centralizar los datos y realizar los informes sin afectar a la base de datos productiva.
- Generar informes en base a la herramienta para obtener información para la toma de decisiones.

- *TOAD for MySQL*: Es una herramienta utilizada por los administradores y analistas de datos para realizar procesos tales como: creación y consultas de tablas.
- *MySQL workbench*: Esta herramienta de visualización de diseño de base de datos la cual es utilizada para explorar la base de datos MOODLE en lo que respecta a como están relacionadas las tablas.
- *MySQL*: Es el sistema de gestión de la base de datos de la plataforma MOODLE.
- *PostgreSQL*: Es el sistema de gestión de base de datos relacional donde se implementa el Data Mart.
- *PGADMIN III*: Es una herramienta que se utiliza en la administración de la base de datos PostgreSQL.
- *Pentaho*: Herramienta de inteligencia de negocio para extraer, transformar y carga de los datos de las diferentes tablas de MOODLE que conforman el Data Mart.
- *Qlikview*: Herramienta de visualización que permite el acceso a la información del Data Mart y facilita el análisis de los datos de MOODLE.
- Acceso a los datos. Los datos utilizados para realizar el Data Mart corresponden a la capacitación realizada en el verano del 2018 en la plataforma académica MOODLE – E-campus y fueron alterados para proteger la información de los docentes y estudiantes de la UTP.

3. Materiales de Trabajo

En esta sección se muestra las herramientas y los datos utilizados para la construcción del Data Mart de la plataforma académica MOODLE de la UTP.

4. Metodología para el desarrollo del modelo de datos

Mediante la metodología ágil Beam* (Business Event Analysis & Modeling) realizamos una serie de eventos y acciones para realizar el modelo de datos ver figura 2 [5], [6].

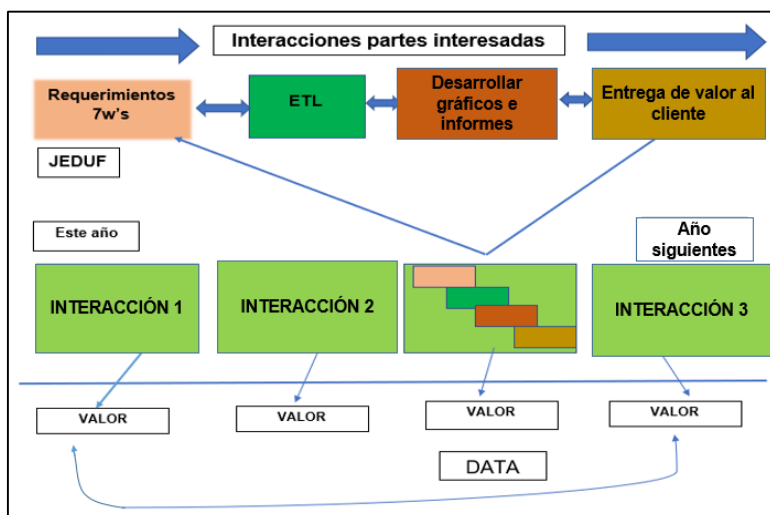


Figura 2. Representación gráfica del modelo BEAM*.

Fuente: Agile Data Warehouse Design. L. Corr and J. Stagnitto.

BEAM* usa las 7W's, *who, what, when, where, how many, why, and how* para descubrir, modelar datos y detalles de los eventos. Cada 7w's se utiliza para construir preguntas o frases que pueden ser utilizadas en las entrevistas con las partes interesadas del negocio, realizar análisis de las fuentes de los

datos, realizar las consultas requeridas para desarrollar las transformaciones de los datos, diseñar el modelo dimensional lógico y físico, realizar las extracciones, desarrollar los tableros de control como resultados y presentarla a los tomadores de decisiones, como se muestra en la figura 3 [5].

7W's	Descripción	Ejemplos
WHO - Quién	Personas y organizaciones	Colaboradores, clientes
WHAT - Qué	Cosas	Productos, servicios
WHEN - Cuándo	Tiempo	Calendario, tiempo del día
WHERE - Dónde	Localización	Tienda, hospital, entregas y direcciones
WHY - Por qué	Razón y casualidad	Promoción, estado del tiempo
HOW MANY - Cuántos	Id de las transacciones y códigos de estados	#ordem, estados de llamadas
HOW - Cómo	Medidas e indicadores	Cantidad, ventas

Figura 3. Ejemplo 7w's para identificar las dimensiones y tablas de hechos.

Fuente: Agile Data Warehouse Design. L. Corr and J. Stagnitto.

5. Resultados

5.1 Implementación primera iteración

Para realizar la primera iteración como solución a la implementación del modelo se realizó los siguientes pasos:

5.1.1 Especificación de requerimientos

Se realizaron entrevista al personal investigador del Centro de Investigación CIDITIC con la finalidad de encontrar respuestas a las 7w's y poder realizar el modelamiento lógico y físico de los datos. Los informes solicitados en la primera interacción son:

- Cantidad de eventos realizados en determinado periodo (año, semestre, mes) clasificado por tipo de acción realizada.

- Total, de eventos por semestre
- Cantidad de eventos realizados clasificando el IP de conexión.
- Top de usuarios que utilizan la plataforma.

5.1.2 Resultados de entrevista identificando las 7w's

Como resultado de las entrevistas (ver figura 4), con las partes interesadas surgieron preguntas como:

- ¿Quién utiliza la aplicación?
- ¿El propósito de utilizar la aplicación?
- ¿Dónde o desde dónde se accede al aplicativo?

7W's	Descripción
WHO - Quién	Usuarios
WHAT - Qué	Cosas
WHEN - Cuándo	Año, semestre, trimestre, mes, día, hora
WHERE - Dónde	Clasificación de IP
WHY - Por qué	No se identificó
HOW MANY - Cuántos	No se identificó
HOW - Cómo	No se identificó

Figura 4. 7w's identificadas para modelar las dimensiones y tablas de hechos para la primera iteración.

Fuente: Elaboración propia.

5.1.3 Modelo lógico de datos

Como resultado de la entrevista al personal investigador del Centro CIDITIC y el análisis realizado en la estructura de base

de datos de la plataforma académica se diseñó el modelo lógico para esta primera iteración ver figura 5.

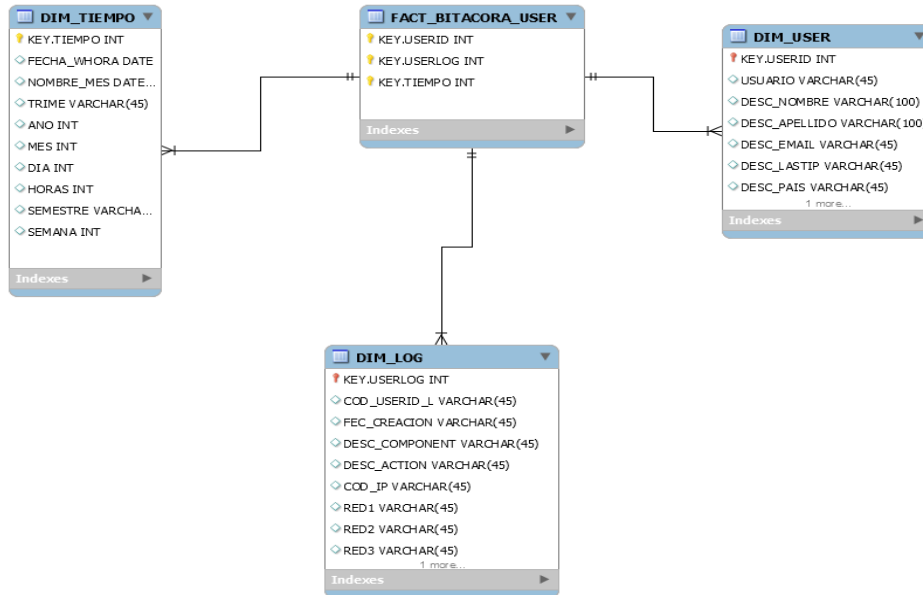


Figura 5. Modelo lógico de datos primera interacción.

Fuente: Elaboración propia.

5.1.4 Desarrollo del ETL – Extraer – Transformar y cargar

Para este proceso se utilizó la herramienta de código libre *Pentaho* (KETTLE), herramienta utilizada para automatizar y monitorear los procesos de ETL y el motor de base de datos utilizado para almacenar los datos fue *POSTGRESQL*.

Se identificó como fuente de datos en la base de datos de la plataforma MOODLE las siguientes tablas: MDL_USER, MDL_LOG_STANDARD.

Para cargar el modelo lógico de datos, se realizaron transformaciones o extracciones de información por medio de consultas SQL y a su vez un mapeo de datos para garantizar la integridad de la información. La Dimensión de tiempo y la tabla de hechos se obtiene de los datos previamente cargados en la dimensión del *log* del repositorio central de *PostgreSQL* (ver figura 6).

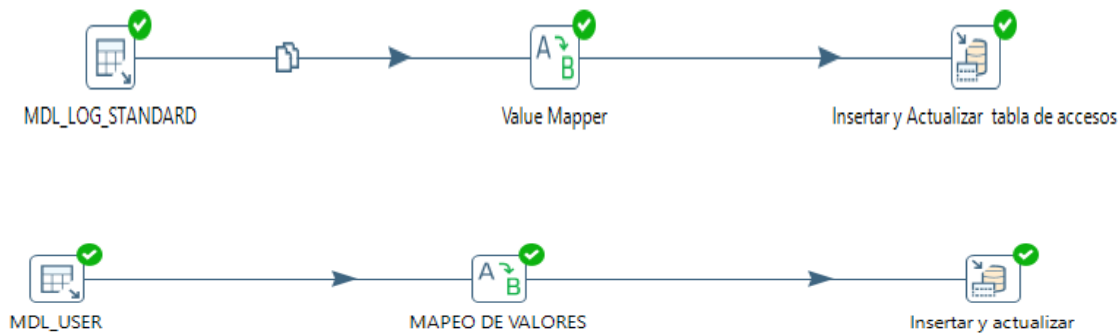


Figura 6. Ejemplo de transformación realizada en *Pentaho* (KETTLE) de la tabla de usuarios y la tabla MDL_LOG_STANDARD de la plataforma académica MOODLE.

Fuente: Elaboración propia.

Estas transformaciones fueron incluidas en el *job* de la herramienta de *Pentaho* (KETTLE) con el propósito que se ejecute en un orden específico y de forma automática la

actualización de nuestro repositorio de datos en *PostgreSQL* (ver figura 7).

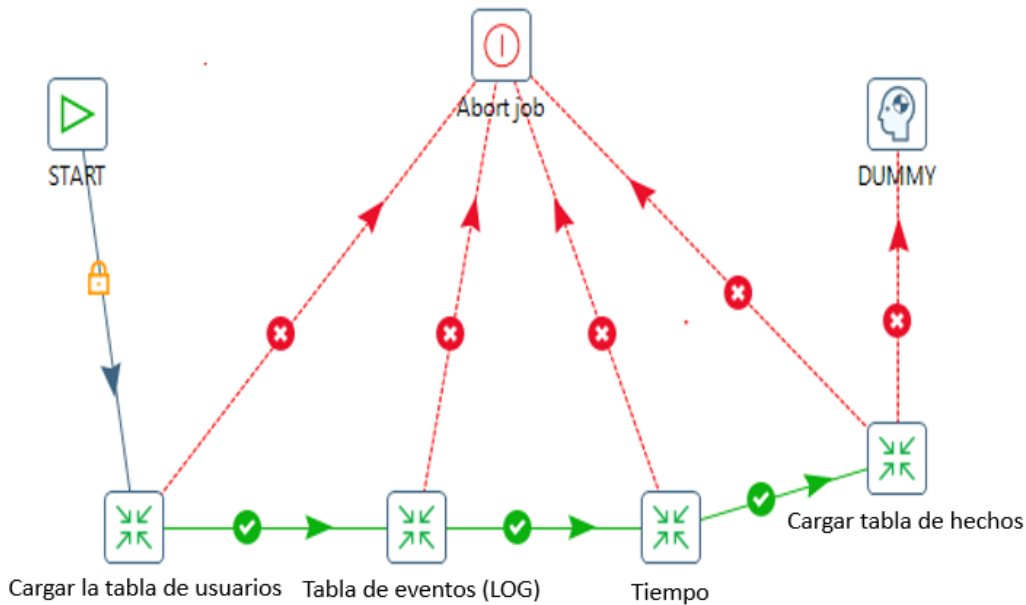


Figura 7. Transformaciones de carga de datos para las tablas de dimensiones y tabla de hechos que corresponde a los procesos requeridos para realizar la primera entrega de valor al cliente.
Fuente: Elaboración propia.

5.1.5 Visualización de datos entrega de valor al cliente

La visualización de datos es una representación gráfica de información y datos.

La herramienta utilizada para la visualización de datos es *Qlikview*, la cual es una herramienta de visualización que permite el acceso a diferentes fuentes de datos y facilita el análisis de los datos. Sus tipos de licencia son: *Personal*

Edition con acceso ilimitado, *Enterprise Edition* y Licencias gratis para docentes, estudiantes e investigadores.

- **Presentación inicial**

Aquí se muestra la aplicación que incluye las fechas de creación, última modificación y recarga del modelo (ver figura 8).



Figura 8. Visualización de la pantalla inicial de la aplicación de *Qlikview*.
Fuente: Elaboración propia.

5.1.6 Tablero de Mando o Dashboard

Aquí se establecen las métricas como resultado del análisis de los requerimientos de los usuarios y así los decisores puedan tomar decisiones, oportunas y estratégicas.

Las mediciones incluidas en el tablero de mando son: clasificación de los eventos por dirección de IP, para realizar esta clasificación, el Centro de Investigación indicó método para obtener la información solicitada. Adicional se incluyeron cantidades de eventos por acción, eventos por un determinado periodo entre otros (ver figura 9).

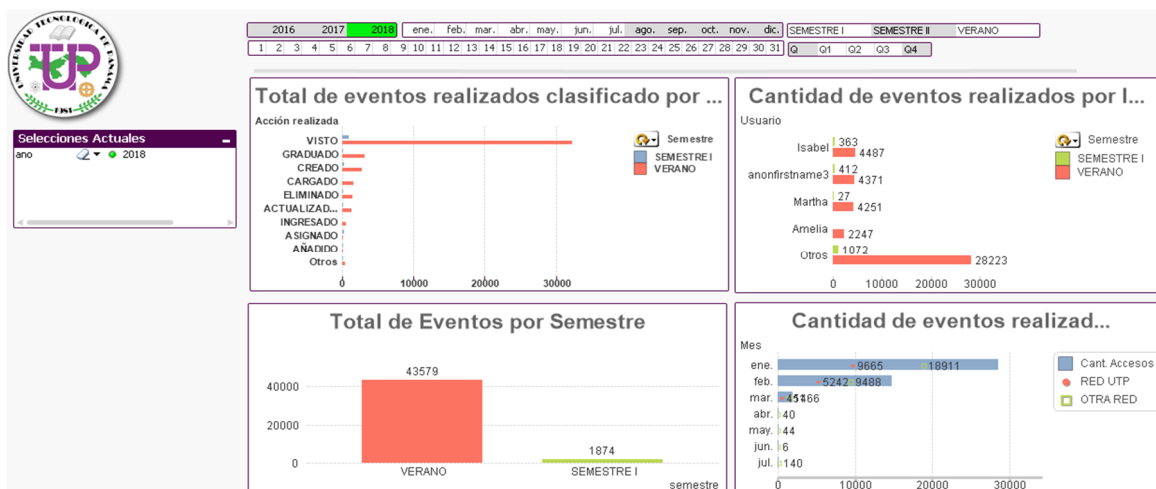


Figura 9. Tablero de mando de la primera entrega de valor al cliente, se abarcan cuatro gráficos.
Fuente: Elaboración propia.

5.1.7 Informes

Se presentan los datos en valores numéricos con el propósito de ser utilizado para los análisis de información al detalle para realizar otras mediciones no establecidas que un personal operativo de la UTP así lo requiere (ver figura 10).

Se adicionaron los siguientes informes:

- Cantidad de eventos realizados por los usuarios en un determinado periodo.
- Cantidad de eventos clasificados por tipo de acción y periodo.
- Cantidad de eventos por mes clasificados por IP.

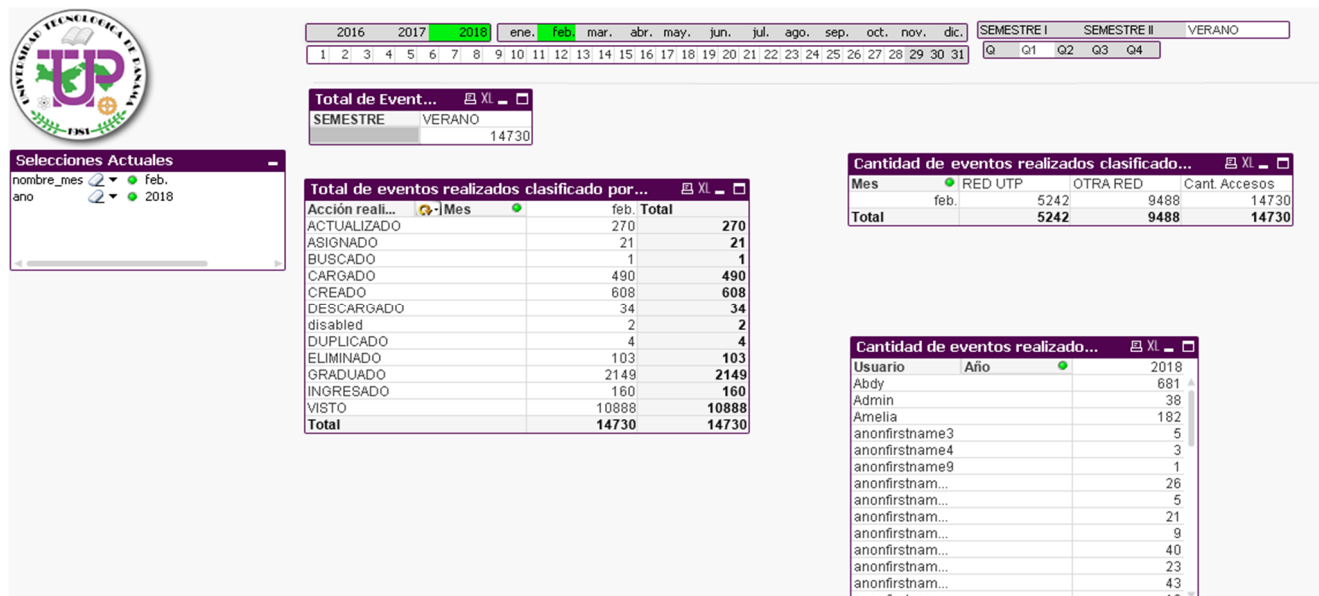


Figura 10. En esta pestaña mostramos los informes en forma de cuadros de tablas.
Fuente: Elaboración propia.

5.2 Resultados - Implementación segunda iteración

5.2.1 Especificación de requerimientos

Se realizó entrevista al personal investigador del Centro CIDITIC con la finalidad de encontrar respuestas a las 7w's (ver figura 11), y poder realizar el modelamiento lógico y físico de los datos para la segunda iteración. Los informes solicitados son los siguientes: cantidades de usuarios por ciudad, país y rol. Además, se solicitó informes por eventos, hora y grupo.

5.2.2 Resultados de entrevista identificando las 7w's

Como resultados de las entrevistas surgen las siguientes preguntas:

- ¿A qué grupos pertenecen los usuarios?
- ¿Qué rol tienen los usuarios?

7W's	Descripción
WHO - Quién	Usuarios
WHAT - Qué	Acción (visto, enviado, enviado), rol, grupos
WHEN - Cuándo	Año, semestre, trimestre, mes, día, hora
WHERE - Dónde	Clasificación de IP, ciudad, país
WHY - Por qué	No se identificó
HOW MANY - Cuántos	eventos realizados en un determinado periodo, eventos por usuarios.
HOW - Cómo	No se identificó

Figura 11. 7w's identificadas para modelar las dimensiones y tablas de hechos para la segunda iteración.

Fuente: Elaboración propia.

5.2.3 Modelo de datos lógicos

Como resultado de la segunda entrevista realizada a las partes interesadas se definió el siguiente modelo de datos (ver figura 12).

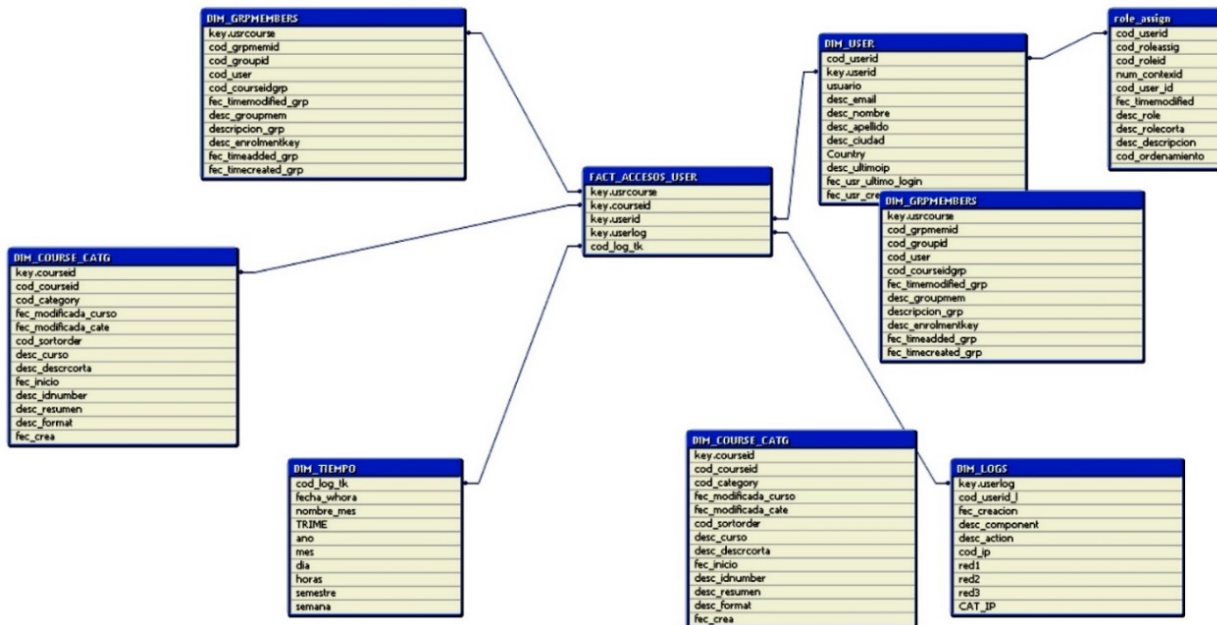


Figura 12. Modelo lógico de datos segunda iteración.

Fuente: Elaboración propia.

5.2.4 Desarrollo del ETL

Para la segunda iteración utilizamos el procedimiento utilizado en la iteración 1 y se le adicionó las transformaciones de grupo y rol para la iteración 2.

El job que se muestra en la figura 13, el recorrido en color verde detalla la ejecución satisfactoria de cada una de las transformaciones del modelo para actualizar el repositorio de PostgreSQL donde se encuentra el modelo de datos.

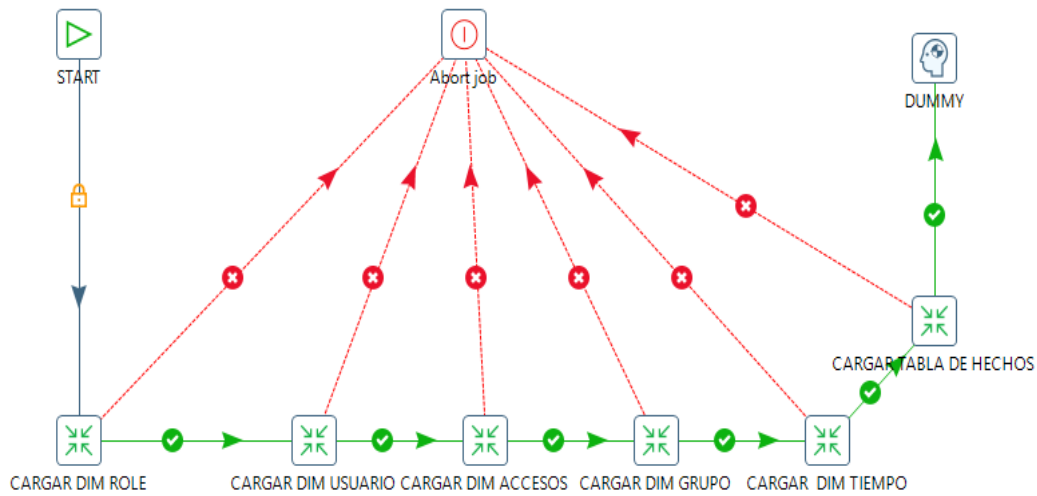


Figura 13. Ejemplo de transformación realizada en *Pentaho* (Spoon.bat) de la tabla de usuarios y la tabla MDL_ROLE de la plataforma académica MOODLE.
Fuente: Elaboración propia.

5.2.5 Visualización de datos entrega de valor al cliente

Se reutiliza el modelo inicial y se añaden nuevas mediciones que generan valor a los altos directivos de la UTP, como se muestra en la figura 14 del tablero de mando o *Dashboard* donde se adicionan otros gráficos, tales como: cantidad de usuarios registrados clasificados por rol, clasificación de usuarios por ciudad y país en donde se puede

observar que esos datos no están siendo ingresados en la plataforma y la universidad requiere evaluar cual es la importancia de obtener ese dato y realizar planes de acciones para que esos datos sean ingresados en el momento de crear los usuarios. Además, podemos generar otros informes tipo numéricos en forma de tablas como se muestra en la figura 15.

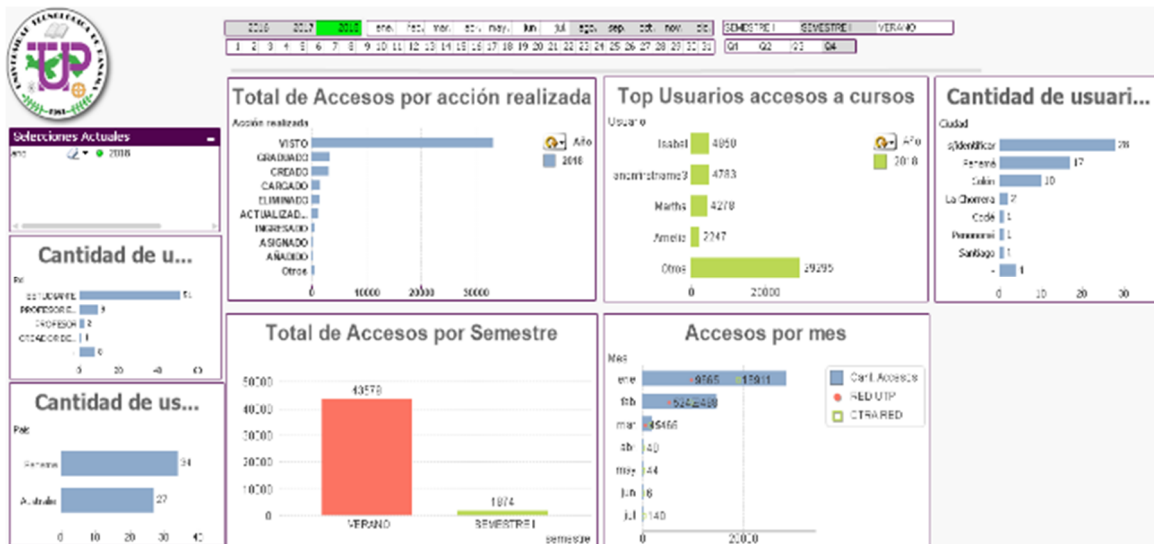


Figura 14. Tablero de mando primera entrega de valor al cliente.
Fuente: Elaboración propia.

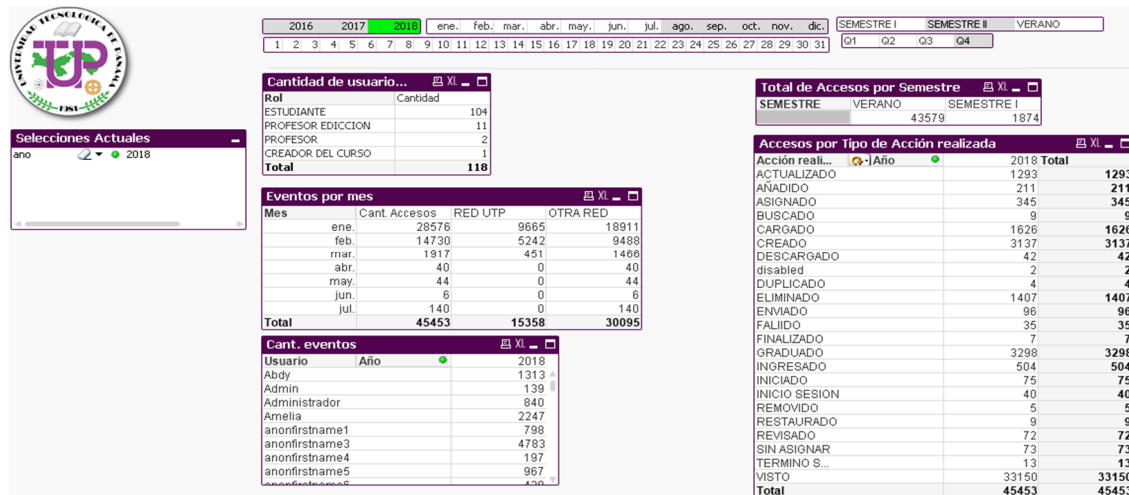


Figura 15. Informes en forma de tablas.
Fuente: Elaboración propia.

6. Conclusiones y recomendaciones

La UTP está realizando esfuerzos para alcanzar la meta de ser la Universidad de referencia a nivel nacional y regional en el tema de educación virtual. Por consiguiente, una solución de inteligencia de negocios aplicada a la plataforma académica MOODLE, permitirá obtener información precisa y oportuna que sirvan de apoyo en la toma de decisiones y la búsqueda del mejoramiento de la calidad en la gestión y la acreditación universitaria, basadas en antecedentes históricos o de lo que está sucediendo en el periodo vigente.

La inteligencia de negocios es una solución que debe ser utilizada en diferentes sectores tales como: académico, salud, instituciones gubernamentales, banca, financieras, entre otras, considerando que los datos hoy en día son los activos más valiosos que tienen las organizaciones convirtiendo los datos para transformarlos en información estratégica para la toma de decisiones.

Con la implementación de un Data Mart, como solución de inteligencia de negocios, se mejora los costos de los procesos administrativos de la UTP, existirá un aumento de satisfacción de los usuarios al tener procesos de negocios automatizados para generar indicadores y gráficos de manera más rápida, siempre y cuando se seleccionen las herramientas de inteligencia de negocio para generar el modelo lógico de datos utilizando una metodología que nos guíe para la implementación.

AGRADECIMIENTO

Damos gracias al Centro de Investigación CIDITIC por habernos facilitado los datos para el desarrollo del Data Mart.

Al investigador Kexy Rodríguez por sus explicaciones en relación a la plataforma MOODLE.

REFERENCIAS

- [1] A. Sanchez, "La agilidad de los negocios es la agilidad del Business Intelligence | Blog Data IQ," 2013. [Online]. Available: <https://dataiq.com.ar/blog/la-agilidad-del-business-intelligence/>.
- [2] Detecsys Soluciones TI, "Qué es un Entorno Virtual de Aprendizaje (EVA)," 2015. [Online]. Available: <https://www.aula1.com/entorno-virtual-aprendizaje-eva/>.
- [3] CIDITIC, "Ecosistema Virtual de Aprendizaje para la universidad Tecnológica de Panamá," 2016.
- [4] Moodle.org, "Moodle - Open-source learning platform," 22 06 2018. [Online]. Available: <https://moodle.org/>.
- [5] L. Corr and J. Stagnitto, Agile Data Warehouse Design, DECISION PRESS, 2014.
- [6] M. MUNTEAN and T. SURCEL, "Agile BI – The Future of BI," *Informatica Economică*, vol. 17, 3 2013.