

## Sistema de ayuda a la decisión basado en un data warehouse

*Carmen Rojas-Muñoz<sup>1</sup>, Víctor Saquicela<sup>2</sup>*

<sup>1</sup> Dirección de Tecnologías de Información y Comunicación, Universidad de Cuenca, Av. 12 de Abril s/n y Agustín Cueva, Cuenca, Ecuador, 010150.

<sup>2</sup> Departamento de Ciencias de la Computación, Universidad de Cuenca, Av. 12 de Abril s/n y Agustín Cueva, Cuenca, Ecuador, 010150.

Autores para correspondencia: {carmita.rojas, victor.saquicela}@ucuenca.edu.ec

Fecha de recepción: 4 de junio 2017 - Fecha de aceptación: 20 de agosto 2017

### RESUMEN

Actualmente la Universidad de Cuenca opera varios sistemas informáticos que generan grandes cantidades de datos, esto representa una heterogeneidad desde diferentes puntos de vista tales como: base de datos, lenguajes de programación, esquemas, datos, etc. Esta heterogeneidad ha generado que cada sistema informático funcione de manera autónoma, ocasionando que el acceso a los datos de manera integrada se convierta en un cuello de botella, puesto que cada sistema posee su propio listado de informes de manera aislada. Por esto, emitir informes integrados a partir de los sistemas se ha convertido en todo un reto para la Universidad. En este trabajo se realizó la construcción de un data warehouse donde se integraron los datos de todos los sistemas informáticos para poder emitir informes consolidados y a futuro poder extraer conocimiento. Para lograr esta integración se combinaron tres herramientas: Hefesto para la construcción del data warehouse, metodología ágil (SCRUM) para el desarrollo del proyecto y BPMN para el modelamiento de procesos ETL.

Palabras clave: Data warehouse, *datamart*, Hefesto, decisiones, indicadores, educación superior.

### ABSTRACT

Today the University of Cuenca operates several information systems each one generating large amounts of data, representing heterogeneity from different points of view such as: database, programming language, schema, data, etc. The heterogeneity is cause that each information system operates stand-alone, causing a variety of problems when all systems access simultaneously the data. Furthermore, each system has its own list of reports, and deriving an integrated report is a huge challenge. To remedy the problem a data warehouse was set-up, allowing that all the data from the different computer systems are integrated enabling the production of complete historic and future reports. To achieve the integration, the research team used the following three tools: Hefesto to create the data warehouse, Agile methodology (SCRUM) for the project development and BPMN for the ETL process modeling.

Keywords: Data warehouse, *datamart*, Hefesto, decisions, indicators, higher education.

## 1. INTRODUCCIÓN

Hoy en día, la Universidad de Cuenca cuenta con 16 sistemas de información, los mismos que día a día generan gran cantidad de información, esta se almacena en las bases de datos de los sistemas: académicos, de investigación, y de gestión con los que cuenta la Universidad de Cuenca. Sin embargo, el gran problema radica en la falta de integración de los sistemas, puesto que no existe una base de datos única y centralizada que permita obtener fácilmente la información de forma rápida y confiable. La información es necesaria para que las autoridades puedan tomar decisiones, para obtener los indicadores establecidos por los organismos de gobierno y para realizar una autoevaluación de la calidad de la

institución, permitiendo establecer sus logros, fortalezas, dificultades y debilidades. Por lo antes mencionado, es necesario que la Universidad de Cuenca implemente una herramienta para integrar, depurar y procesar los datos, con el objetivo de poder realizar análisis de la información desde diferentes puntos de vista. Esta herramienta es ideal para la toma de decisiones en cualquier área de la institución, basadas en información integrada y global y principalmente con el objetivo de poder obtener la información necesaria para todos los indicadores solicitados por los organismos de control para un proceso de evaluación y acreditación. Se planteó la elaboración de un data warehouse para la Universidad, el mismo que fue implementado en base a un profundo análisis de las necesidades de cada una de las áreas de la institución. El data warehouse integra toda la información y está disponible en línea para todas las autoridades y quienes trabajan diariamente con la información para la toma de decisiones. A través de este trabajo se definió el modelo para la construcción de un sistema de ayuda a la toma de decisiones que ha servido para la Universidad de Cuenca y podría ser replicado por otras o incluso por todas las Universidades del país (Ecuador).

El resto del trabajo está organizado de la siguiente manera: la sección 2 describe los antecedentes referentes a los conceptos de data warehouse, y también se describe casos similares en Latinoamérica, en Europa y en los Estados Unidos; la sección 3 detalla cómo se define la metodología utilizada para la construcción de data warehouse y en qué consiste cada componente que hace parte de la arquitectura definida. Finalmente, en la sección 4 se describen algunas conclusiones sobre el desarrollo del trabajo y los trabajos futuros.

## 2. ANTECEDENTES Y TRABAJOS RELACIONADOS

En esta sección se describen los conceptos principales sobre las metodologías para el desarrollo de un data warehouse y se analizan algunos trabajos relacionados que guiarán la descripción del trabajo.

### 2.1. Antecedentes

Para el desarrollo de un data warehouse, existen algunas metodologías, entre ellas están: Inmon, Kimball y Hefesto. La metodología de Bill Inmon señala que se debe partir de un data warehouse para luego llegar a los *datamarts*, es decir será una metodología *top-down* (Inmon, 2002); por otro lado, Ralph Kimball propone su metodología de forma contraria a la de Inmon, es decir, parte de los *datamarts* para llegar a construir el data warehouse (Ross & Kimball, 2013), siendo entonces una metodología *bottom-up*. Por otro lado, la metodología Hefesto, la cual se escogió para el desarrollo de este trabajo, plantea una combinación de las dos metodologías (Inmon y Kimball), lo cual permite implementar el data warehouse a partir de requerimientos de los usuarios para construir un esquema lógico y la definición de los procesos ETL (Bernabeu, 2010).

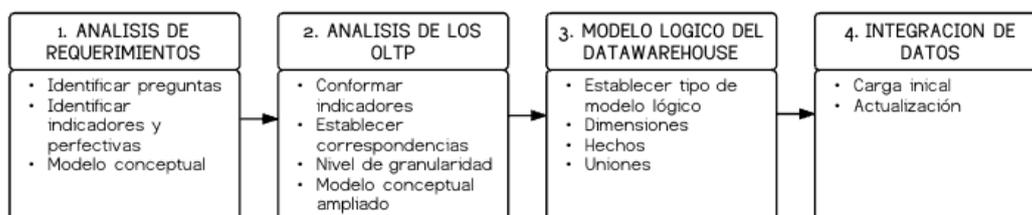
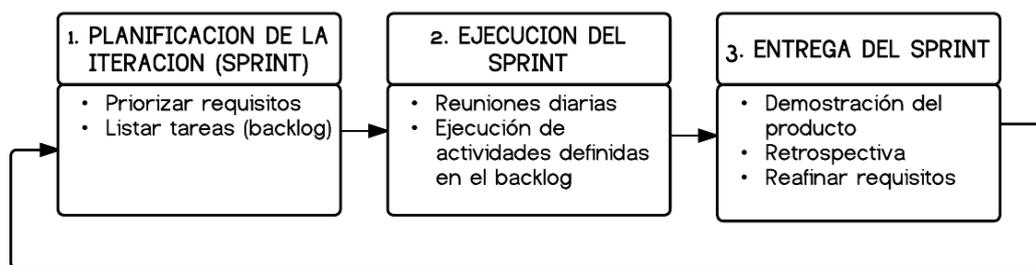


Figura 1. Metodología Hefesto.

La Figura 1, indica gráficamente cada uno de los pasos de la metodología Hefesto. El primer paso de Hefesto da como resultado las preguntas claves para los usuarios y el modelo conceptual del data warehouse. En el segundo paso se analizan las fuentes OLTP para establecer cómo se calculan los indicadores y las correspondencias con el modelo lógico definido con cada uno de los campos que se incluirían en cada perspectiva. El tercer paso parte del modelo conceptual para definir el tipo de modelo que se utilizara y se definen las tablas de hechos y de dimensiones y las respectivas uniones entre las

tablas. El paso final de Hefesto consiste en poblar con los datos las tablas definidas en el modelo lógico, a través de un proceso ETL (extracción, transformación y carga), de manera que en el data warehouse se almacenen los datos sin errores.

Al hablar del desarrollo de proyectos de software, existen varias metodologías, sin embargo, la metodología SCRUM se destaca por ser una metodología ágil y que permite interactuar con el cliente y entregar resultados parciales de manera que éste se sienta satisfecho y comparta sus inquietudes, e incluso para solicitar cambios durante el desarrollo del proyecto. SCRUM también permite establecer iteraciones cortas de desarrollo, denominadas *sprints*. Se planifica cada *sprint* y, durante su desarrollo, se realizan reuniones diarias para revisar el avance de las tareas. Al final de cada iteración se realiza una retrospectiva y la entrega al cliente de manera que pueda revisar los resultados del proyecto. En la Figura 2, se visualiza cómo funciona la metodología SCRUM. En la planificación de cada iteración se priorizan los requisitos, lo que da como resultado una lista de tareas. En la ejecución misma del *sprint* se dan reuniones diarias entre los miembros del equipo y se ejecutan cada una de las tareas definidas, por último, se realiza la entrega del producto resultado de la iteración y los clientes pueden refinar los requisitos para mejorar el producto presentado.



**Figura 2.** Metodología ágil de desarrollo de proyectos SCRUM.

Con respecto a los procesos del negocio, es importante que una empresa pueda modelar sus procesos, los mismos que son una colección de actividades o tareas que están relacionadas para lograr un objetivo y producir un servicio o producto del negocio. En este trabajo para modelar los procesos ETL (extracción, transformación y carga), se utilizó BPMN, puesto que provee una notación gráfica para definir y entender un proceso de negocio de la empresa de manera fácil, entendible y estándar.

Basado en lo descrito anteriormente, un reto que se presenta es la integración de Hefesto, SCRUM y BPMN, lo que permitirá generar un proceso (BPMN) unificado de creación de un data warehouse (Hefesto) de manera ágil (SCRUM) que permita a futuro ser reproducido por otras universidades.

## 2.2. Trabajos relacionados

Se analizaron varios casos sobre el uso de un data warehouse en una empresa, concluyendo que este tipo de soluciones son útiles para la toma de decisiones. Se pudo observar que un data warehouse puede ser aplicado en cualquier empresa no importa cuál sea su actividad. Este tipo de sistemas existen en empresas de actividad financiera con el fin de examinar a sus deudores y cartera vencida (Lozada & Cruz, 2014), o para empresas sanitarias, en donde se considera fundamental dotar de un sistema para la toma de decisiones para los directivos (Castillo, 2014). En todos los casos analizados, el resultado ha sido de éxito, puesto que se eliminaron tareas manuales y se tiene disponibilidad en línea de los datos a través de los informes y cuadros de mando generados para los usuarios.

Con respecto a la aplicación de este tipo de sistemas en la educación y principalmente en la educación superior, se encontraron algunos casos como por ejemplo la implementación de sistemas para la toma de decisiones en el ámbito universitario de Argentina (Menéndez & De Lujan, 2012), este proyecto señala que trabajar con un data warehouse simplifica procesos para tomar decisiones, puesto que el acceso es fácil y es beneficioso para el usuario que pueda hacer sus propias consultas. Un análisis del rol del datawarehousing en la educación superior fue realizado por (Guan, Nunez & Welsh, 2002), en este artículo se examinaron las metas y retos de las universidades con respecto al manejo de la

información, necesaria para la toma de decisiones y para planificar estratégicamente. El mencionado trabajo plantea que un data warehouse puede servir como una herramienta para la gestión del conocimiento en el área académica, puesto que almacenará gran cantidad de información útil y necesaria para el soporte de la toma de decisiones. Por otro lado, una investigación realizada en la Universidad de Phoenix por Pirani & Albertch (2015) indicó que todas las Universidades buscan tener una herramienta que les permita realizar el análisis de información relevante que le permita tomar decisiones acertadas. Los autores concluyeron que la institución puede crear sus propios indicadores y evaluarlos a través de una herramienta amigable sobre la cual deben estar bien capacitados para que sea de gran utilidad.

Por todo lo antes mencionado, resulta evidente que un data warehouse hoy en día es de gran ayuda para la toma de decisiones de los más altos directivos de una empresa. En este sentido, el desarrollo e implementación del data warehouse para la Universidad de Cuenca permitirá también obtener los indicadores necesarios para que las autoridades de la Institución puedan tomar decisiones en base a información segura, oportuna y confiable.

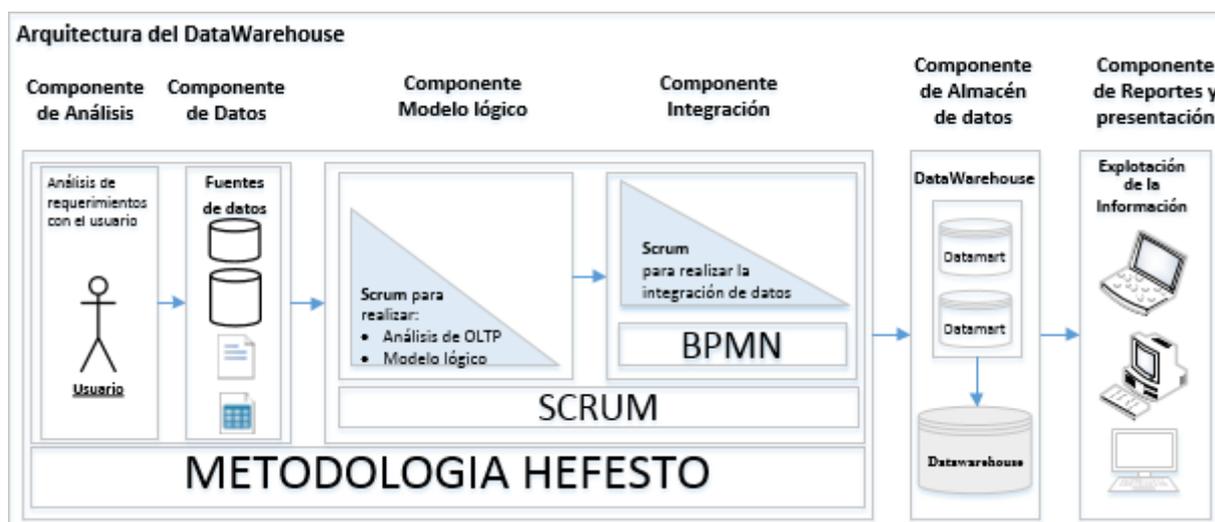
### 3. PROCESO DE CONSTRUCCIÓN E IMPLEMENTACIÓN DEL DATA WAREHOUSE

En esta sección se describe la arquitectura propuesta para el desarrollo de un data warehouse, para ello se utiliza un ejemplo práctico sobre un *datamart* específico. Este *datamart* guiará todo el proceso propuesto.

#### 3.1. *Arquitectura del data warehouse*

La arquitectura planteada en este trabajo para la construcción del data warehouse y para cualquier *datamart* combina la metodología Hefesto para la construcción de un data warehouse, la metodología ágil de desarrollo de proyectos SCRUM y la notación BPMN para el modelado de los procesos de negocio y específicamente para modelar los ETLs necesarios para el data warehouse. Esta combinación se realizó con el objetivo de obtener un producto que esté acorde a las necesidades del usuario, puesto que, aplicando SCRUM, la participación del usuario con el equipo de desarrollo permitió avanzar y ganar tiempo y con menos esfuerzos, al detectar errores durante la aplicación de la metodología Hefesto para la construcción del data warehouse. Por otro lado, la notación BPMN al ser estándar y conocida permite leer más rápidamente cómo se modelaron cada uno de los procesos ETL para la construcción del data warehouse. Aplicando esta arquitectura se pudo demostrar que el uso de SCRUM juntamente con Hefesto permitió obtener productos mejorados por cada iteración o *sprint* que se ejecutaron durante el desarrollo del trabajo. Además, el uso de BPMN sobre los ETL permiten gestionar de manera efectiva todo el proceso de construcción del data warehouse.

La Figura 3 visualiza la arquitectura definida para la construcción del data warehouse de la Universidad de Cuenca. Como se puede observar en la figura, se combina la metodología Hefesto, la metodología ágil SCRUM para desarrollo de proyectos y la notación BPMN para la construcción del Data Warehouse. Esto se hizo con el objetivo de obtener un producto que esté acorde a las necesidades del usuario, puesto que la participación del usuario con el equipo de desarrollo permite avanzar y ganar tiempo y esfuerzos al detectar los posibles errores durante la aplicación de la metodología Hefesto. Por otro lado, la notación BPMN al ser estándar y conocida permite leer más rápidamente cómo se modeló cada uno de los procesos ETL para la construcción del data warehouse. Con la combinación descrita anteriormente, se pretende demostrar que el uso de SCRUM juntamente con Hefesto permite obtener productos mejorados por cada *sprint* ejecutado, lo cual permite detectar rápidamente errores o cambios en los requerimientos. Además, el uso de las anotaciones BPMN sobre los ETL permiten gestionar de manera efectiva todo el proceso de construcción del data warehouse. Esta arquitectura está conformada por seis componentes, los mismos que se pueden visualizar en la Figura 3 y que se procede a explicar a continuación:



**Figura 3.** Arquitectura del data warehouse.

### Componente de análisis

Este componente pertenece a la primera fase de la metodología Hefesto para la construcción de cada uno de los *datamart*. Estos *datamarts* fueron definidos luego de realizar entrevistas, encuestas a los usuarios de las áreas más importantes de la Universidad de Cuenca. El resultado de este proceso se muestra en la Tabla 1, donde se listan las dependencias analizadas y los *datamarts* con sus respectivos indicadores. La metodología Hefesto inicia identificando los requerimientos de los usuarios, puesto que ellos serán quienes den la pauta de lo que se espera del almacén de datos. Adicional a las entrevistas realizadas con los usuarios, se procedió también a revisar los indicadores solicitados por los organismos de control, el Consejo de Educación Superior (CES)<sup>1</sup> y el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEACES)<sup>2</sup>, para la evaluación y acreditación de las instituciones de nivel superior, dando como resultado un listado de indicadores más representativo que sirvan tanto para la toma de decisiones interna como para justificar los indicadores dentro de un proceso de evaluación de instituciones de educación superior.

**Tabla 1.** *Datamarts* e indicadores establecidos para el data warehouse.

| Dependencias  | Datamart                                  | Indicadores  |
|---|---|--|
| <ul style="list-style-type: none"> <li>• Facultades</li> <li>• Consejo Académico</li> <li>• Direcciones de carrera</li> </ul> | <i>Datamart</i> de Grado o carreras       | Promedios de estudiantes por periodo lectivo y por carrera                   |
|   |   | Número de estudiantes por grupo ofertado y por docente en un periodo lectivo |
|   |   | Número de Matriculados por Carrera en un periodo lectivo                     |
|   |   | Estudiantes nuevos matriculados por carrera en un periodo lectivo            |
| <ul style="list-style-type: none"> <li>• Comisión de Evaluación Interna</li> </ul>  | <i>Datamart</i> de Evaluación del docente | Número de estudiantes graduados por periodo lectivo y por carrera            |
|   |   | Calificación de los docentes por carrera, facultad y proceso de evaluación   |
|   |   | Calificación obtenida por los docentes según el actor y función              |
|   |   | Porcentaje de participación en los procesos de evaluación                    |

<sup>1</sup> <http://www.ces.gob.ec>

<sup>2</sup> <http://www.ceaaces.gob.ec>

|   |                                   |  |
|---|-----------------------------------|--|
| • Dirección de Talento Humano                             | <i>Datamart</i> de Talento Humano | Número de contratos ocasionales por años y mes   |
|   |                                   | Número de servidores en calidad de asesores  |
|   |                                   | Número de vacantes del área administrativa   |
|   |                                   | Número de docentes que han migrado a estudiar en el extranjero                                 |
|   |                                   | Número de servidores desvinculados de la Institución por año y mes                             |
|   |                                   | Número de solicitud de permisos según el motivo, solicitadas por los servidores universitarios |
|   |                                   | Número de sanciones según el régimen disciplinario   |
|   |                                   | Número de docentes según su formación por año  |
|   |                                   | Número de horas asignadas a los docentes por actividad   |
|   |                                   | Total de salarios  |
| • Dirección de Posgrados                                  | <i>Datamart</i> de Posgrados      | Promedios de estudiantes por periodo lectivo y por programa de posgrado                        |
|   |                                   | Número de matriculados por programa de posgrado y por año                                      |
|   |                                   | Número de posgrados aprobados y en ejecución por año   |
| • Dirección de Evaluación<br>• Dirección de Planificación | <i>Datamart</i> de Proyectos      | Número de participantes por proyecto por dependencia y por año                                 |
|   |                                   | Total de presupuesto por proyecto por dependencia y por año                                    |
| • Dirección de Investigación                              | <i>Datamart</i> de Investigación  | Número de publicaciones por docente y por año  |
|   |                                   | Número de proyectos de investigación: presentados y aprobados por año                          |
|   |                                   | Presupuesto de proyectos de investigación por año  |

En el caso particular de este trabajo se describe el *datamart* de evaluación docente, en el cual se identificaron las siguientes preguntas:

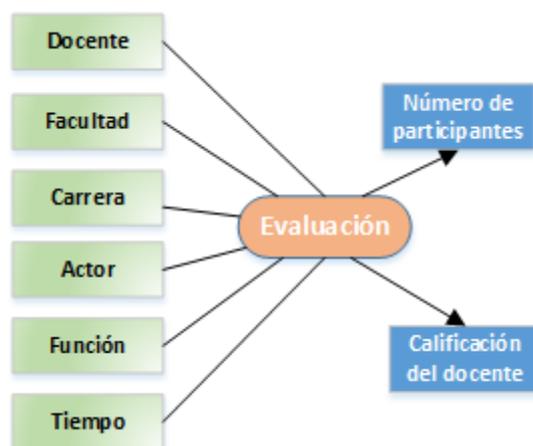
1. Se desea conocer la calificación que han obtenido los docentes durante cada proceso de evaluación, según la facultad y la carrera. En otras palabras “Calificación obtenida por un docente en una carrera y facultad en un proceso de evaluación determinado”.
2. Se desea conocer la calificación que han obtenido los docentes durante cada proceso de evaluación, según el actor y la función. En otras palabras “Calificación obtenida por un docente por parte de un actor en una función, en un proceso de evaluación determinado”.
3. Se desea conocer la participación de los docentes evaluados y evaluadores en cada facultad y carrera, en cada proceso de evaluación. En otras palabras “Número de participantes en el proceso de evaluación, por facultad y carrera en cada proceso de evaluación”.

A partir de las preguntas, se identificaron los indicadores, es decir: los valores numéricos que representan lo que se va a analizar y las respectivas perspectivas utilizadas para los análisis, estos se pueden revisar en la Tabla 2.

**Tabla 2.** Indicadores y perspectivas para el *datamart* de evaluación.

| Indicadores                   | Perspectivas                     |
|-------------------------------|----------------------------------|
| – Calificación                | – Docentes                       |
| – Porcentaje de participación | – Facultades                     |
|                               | – Carreras                       |
|                               | – Actores                        |
|                               | – Funciones                      |
|                               | – Tiempo (proceso de evaluación) |

Una vez que se culminó con el levantamiento de requerimientos y se identificaron las preguntas y por tanto se establecieron los indicadores y las perspectivas de la Tabla 2, se continuó con la elaboración del modelo conceptual; este se lo puede visualizar en la Figura 4.



**Figura4.** Modelo conceptual del *datamart* de evaluación.

#### Componente de datos

En este componente se realiza un análisis de todas las fuentes que están disponibles para la obtención de los datos para cada uno de los posibles *datamarts*. La Universidad de Cuenca almacena los datos en diferentes tipos de bases de datos, entre las cuales están Oracle, DB2 sobre un AS400, Mysql y Postgresql. Además, en algunas dependencias y facultades existe información en archivos de Excel. Para el caso de estudio de este trabajo se detectó que los datos se almacenan en dos bases de datos distintas, la información correspondiente a la academia está en Oracle y la información sobre los docentes está en DB2 sobre un AS400.

Basado en los indicadores y perspectivas definidos en el componente de análisis se realizó un análisis exhaustivo de las bases de datos para identificar si es posible responder a las preguntas planteadas por los usuarios, concluyendo que efectivamente las bases de datos contienen la información suficiente para construir el *datamart* de evaluación docente.

#### Componente modelo lógico

En este componente se inició el uso de la metodología de gestión de proyectos SCRUM, con el objetivo de realizar el análisis de los OLTP y el diseño del modelo lógico, que son el segundo y tercer paso de la metodología Hefesto, respectivamente. Para realizar estas actividades se planificó y ejecutó la primera iteración (*sprint*) con una duración de 15 días, el mismo que fue ejecutado por un grupo de 5 personas.

Se determinó la conformación y manera de cálculo de los indicadores, que fueron establecidos con la ayuda de los usuarios. En la Tabla 3 se puede observar la forma de cálculo de cada uno de ellos.

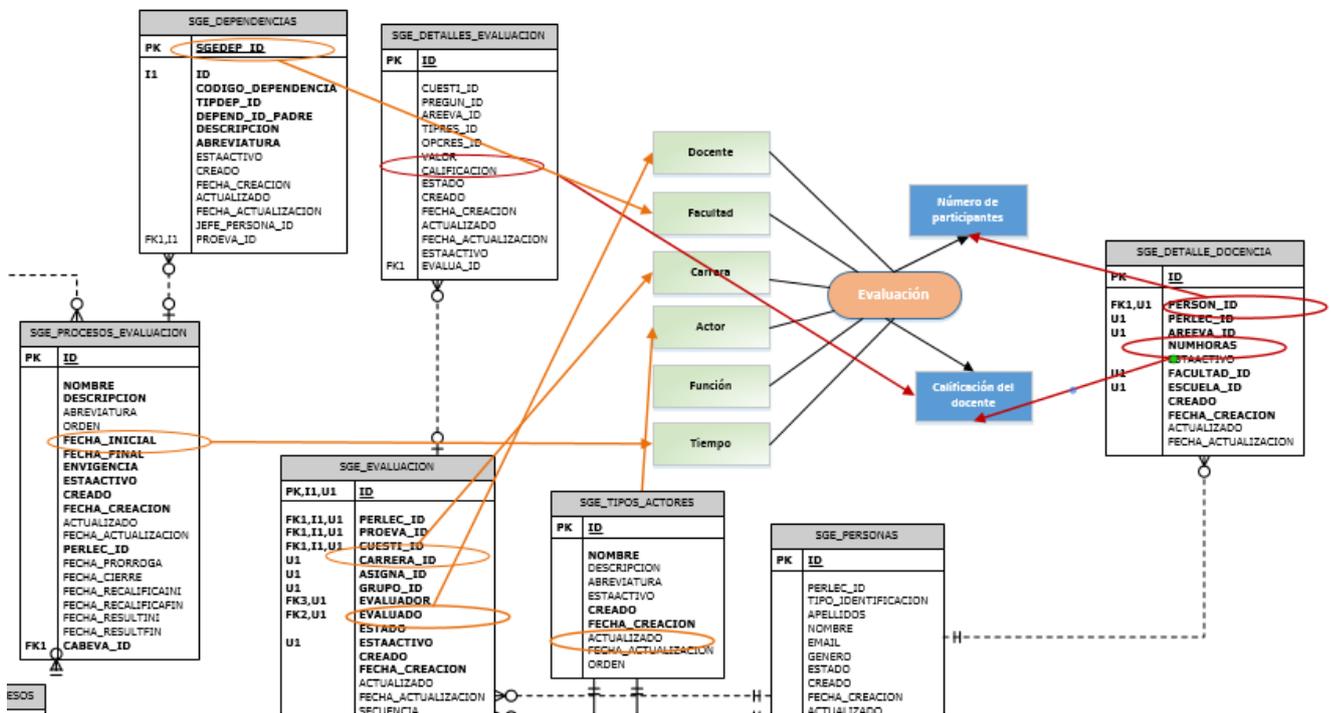
- “Número de participantes”: Este indicador se obtiene de contar el número de actores (evaluadores y evaluados), que intervinieron en el proceso de evaluación.
- “Calificación del docente”: La calificación de los docentes (hecho) se calcula de manera diferente, dependiendo del proceso de evaluación.

**Tabla 3.** Conformación de indicadores para el *datamart* de evaluación.

| Procesos anteriores al 2014                           | Procesos a partir del 2014   |
|---|--|
| $\sum$ Calificaciones del docente según tipo de actor | $(D \times H.D) + (I \times H.I) + (D.G. \times H.D.G)$<br><i>Dedicación docente</i> |

D = Docencia  
H.D = Total de horas de docencia asignadas al docente en su distributivo  
I = Investigación  
H.I = Total de horas de investigación asignadas al docente en su distributivo  
D.G = Gestión académica  
H.D.G = Total de horas de gestión y dirección asignadas al docente en su distributivo  
Dedicación docente: Puede ser Tiempo Completo = 40 horas, Medio tiempo = 20 horas y Tiempo parcial de acuerdo al número total de horas a la semana del docente en su distributivo

El siguiente paso para el desarrollo de este trabajo fue establecer las correspondencias entre las fuentes de datos y el modelo conceptual, tal como se puede observar en la Figura 5.



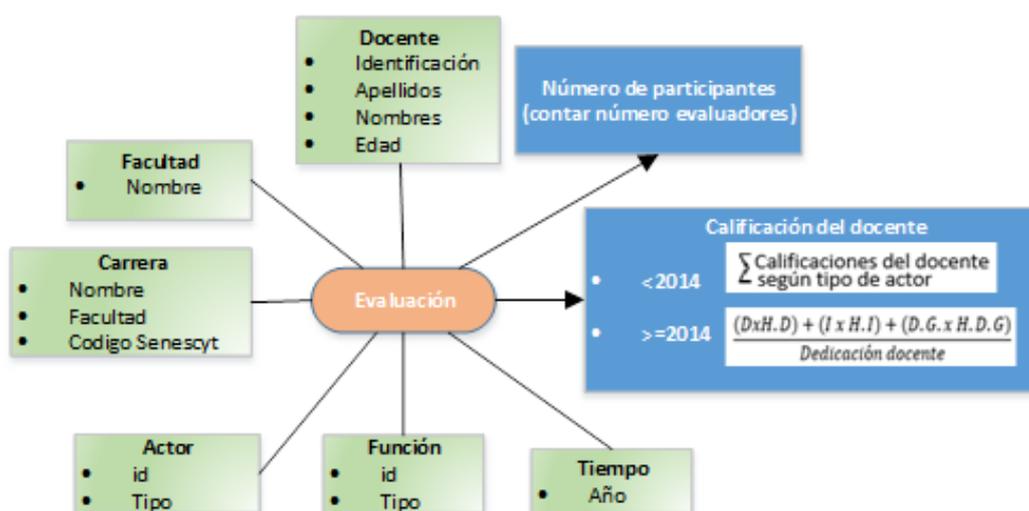
**Figura 5.** Correspondencias entre el modelo conceptual y las fuentes de datos para el *datamart* de evaluación.

La siguiente actividad realizada, correspondiente al segundo paso de Hefesto fue establecer el nivel de granularidad, esto es, definir los campos más relevantes para cada perspectiva, que fueron utilizados para filtrar los indicadores. En el caso específico de la evaluación del docente, primero se analizó el diccionario entidad-relación y el diccionario de datos para definir los campos de cada perspectiva. Como ejemplo en la Tabla 4, se puede visualizar cómo se concibieron los campos para las perspectivas docente y facultad. Además, el ámbito de la perspectiva tiempo se definió que es anual, puesto que un proceso de evaluación en la Universidad se lo realiza cada año lectivo.

**Tabla 4.** Nivel de granularidad de las perspectivas docente y facultad para el *datamart* de evaluación.

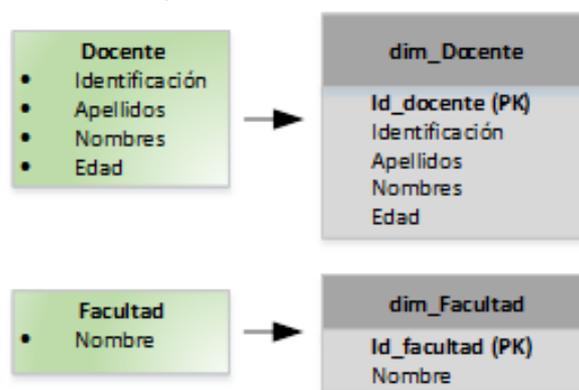
| Perspectiva | Campos  | Diagrama entidad-relación  |
|-------------|---|----------------------------|
| Docente     | <ul style="list-style-type: none"> <li>- person-id</li> <li>- apellidos</li> <li>- nombres</li> <li>- edad</li> </ul> | adminuc.personas_naturales |
| Facultad    | <ul style="list-style-type: none"> <li>- descripción</li> </ul>   | adminuc.dependencias       |

Luego de analizados los campos de las perspectivas, el siguiente paso realizado en el *sprint* ejecutado para el componente modelo lógico, fue establecer el modelo conceptual ampliado, indicando gráficamente los campos de cada perspectiva y la fórmula de cálculo para cada indicador. La Figura 6, muestra cómo se estableció este modelo lógico ampliado para el caso de estudio.

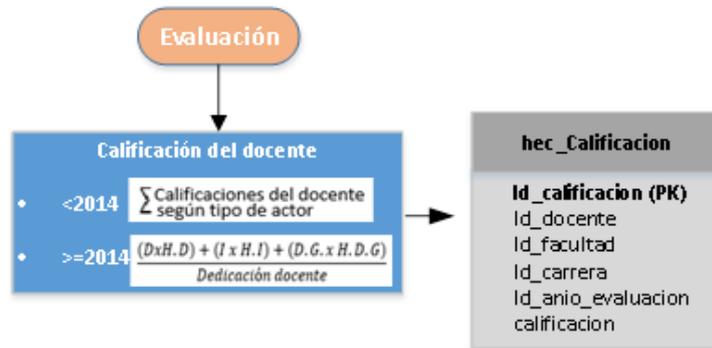


**Figura 6.** Modelo conceptual ampliado para el *datamart* de evaluación.

Para finalizar el *sprint*, la siguiente actividad fue confeccionar el modelo lógico de la estructura del data warehouse, en base al modelo conceptual establecido en el paso anterior. Lo primero que se definió fue el tipo de esquema a utilizar, el mismo que para el caso de evaluación; sería el esquema estrella, por lo que se transformaron las perspectivas en dimensiones y los indicadores en hechos. En las Figuras 7 y 8 se puede visualizar algunas de las tablas de dimensiones y de hechos definidos para el *datamart* que se ha tomado como ejemplo en este trabajo.



**Figura 7.** Ejemplos de tablas de dimensiones para el *datamart* de evaluación.

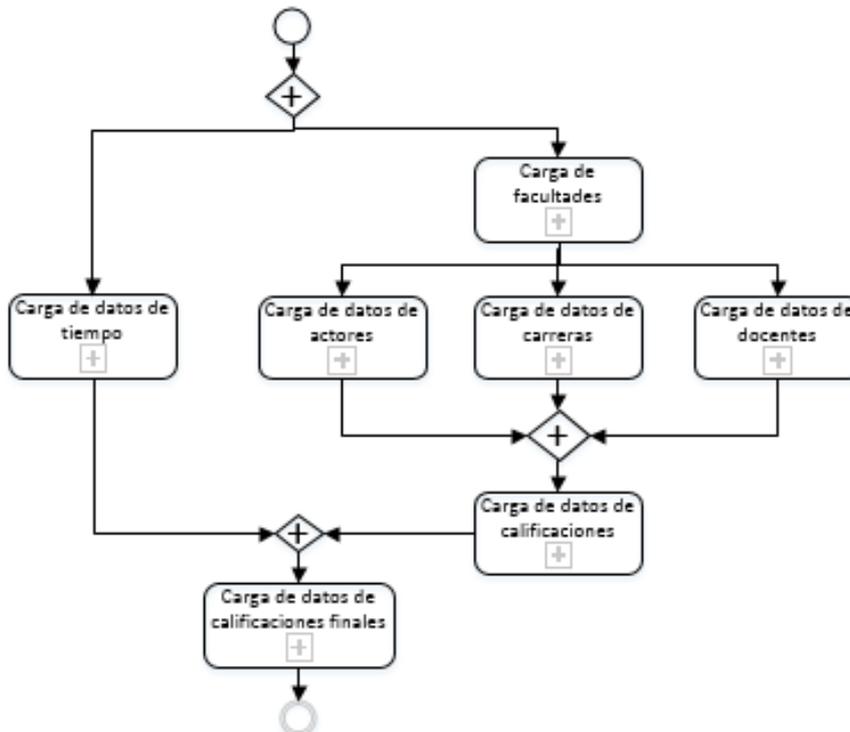


**Figura 8.** Tabla de hechos para obtener la calificación del docente para el *datamart* de evaluación.

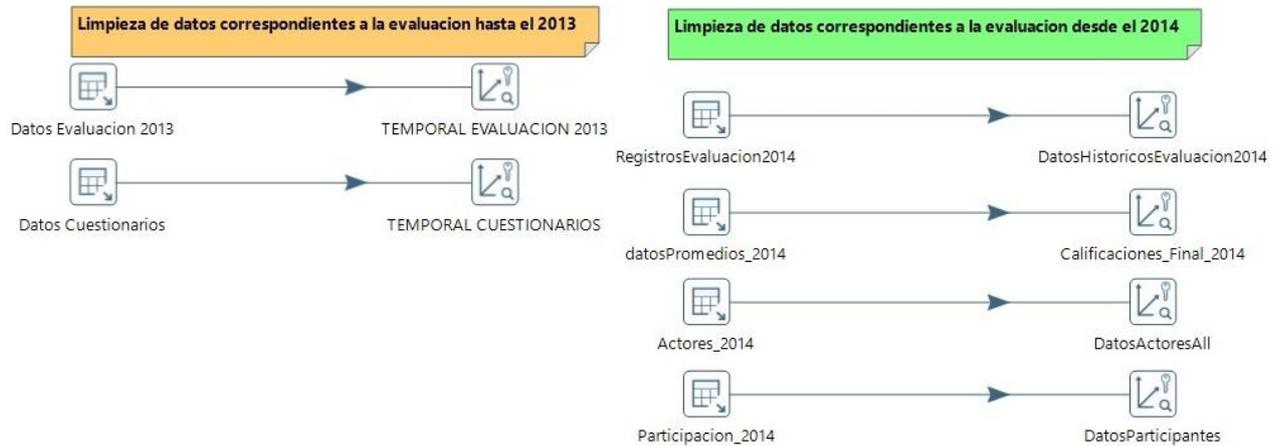
Cada ejecución de un *sprint* genera un producto, que en este caso son los modelos lógicos de datos multidimensionales ampliados con sus respectivas descripciones. Si en una posterior revisión por parte del usuario se detectan errores o inconsistencia, en el nuevo *sprint* se considerarán estos nuevos desafíos a ser considerados lo que permitirá obtener mejores productos.

Componente integración

En el componente de integración se procede a poblar el data warehouse con los datos obtenidos de los sistemas operacionales, sin embargo, los datos deben pasar por un proceso (ETL) de limpieza para evitar datos anómalos. Se continúa con la metodología Hefesto y con el uso de SCRUM, no obstante, se ejecutó una iteración o *sprint* para la construcción de cada proceso ETL, necesarios para la creación del almacén de datos; y es en este punto que se utilizó la notación BPMN, para modelar cada uno de estos procesos ETL. En la Figura 9 se puede ver como se modeló el proceso ETL para la carga de datos de las calificaciones de los docentes utilizando BPMN, y la Figura 10 muestra el proceso ETL que se ejecutó para realizar la limpieza de los datos de las evaluaciones.



**Figura 9.** Modelado con BPMN del proceso ETL para calificaciones.

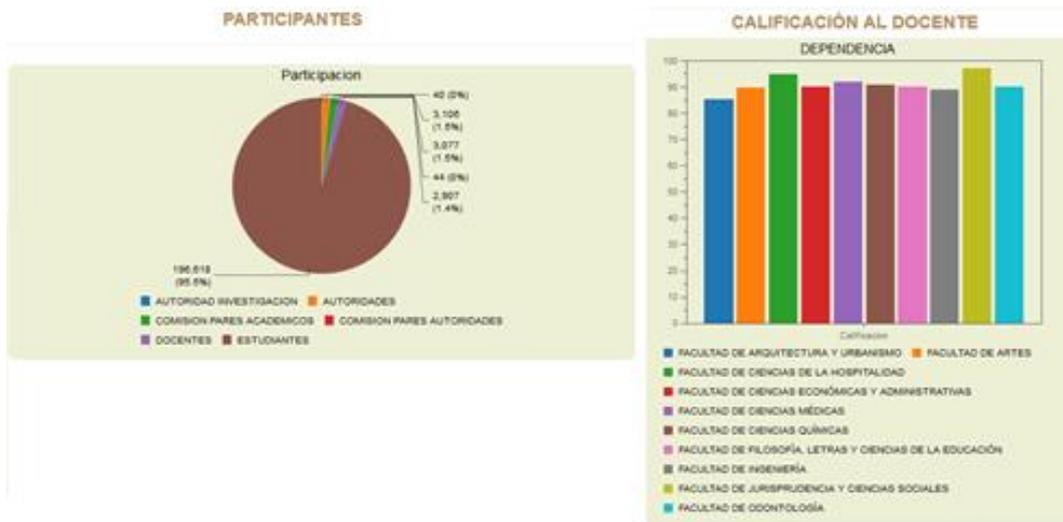


**Figura 10.** Elaboración del proceso ETL para realizar la limpieza de los datos de evaluaciones.

Como paso final de este componente se ejecutó el proceso para poblar las tablas de dimensiones y hechos y se procedió con la creación de los cubos multidimensionales. Los procesos ETL modelados con BPMN se ejecutaron en un *sprint* diferente al mencionado en el componente del modelo lógico, puesto que se está experimentando con dos grupos de personas y se considera que el poblado del modelo puede ser independiente del modelamiento lógico, de todas maneras, a futuro se pretende mezclar estos dos componentes en un solo *sprint*.

Componente de almacén de datos

El componente de almacén de datos se refiere al data warehouse como tal, es decir es el repositorio de los datos que resulta del componente de integración y que forma cada uno de los *datamart* del data warehouse. Cabe indicar que durante todo este trabajo se ha tomado como ejemplo el *datamart* de evaluación. Es importante dar a conocer que el almacenamiento centralizado de datos se realiza utilizando un modelo de data warehouse en estrella sobre una base de datos relacional (Postgresql). En esta base de datos se realiza todas las configuraciones de optimización que permiten a los modelos multidimensionales un rendimiento óptimo.



**Figura 11.** Ejemplos de reportes que se encuentran para el *datamart* de evaluación.

Componente de reportes y presentación

Este componente es el más importante para los usuarios. Es a través de éste que el usuario interactúa con el almacén de datos para obtener sus reportes y datos necesarios para la toma de decisiones. Este

componente fue implementado con la herramienta Pentaho<sup>3</sup>. Para que los usuarios puedan acceder a la herramienta, primero se definió el nivel de acceso a la información creando roles para este efecto y los cubos respectivos. Se crearon usuarios que únicamente puedan acceder a reportes y otros roles para que puedan realizar el análisis de la información. Así también se definió el acceso para cada uno de los *datamarts* implementados.

Este componente también está considerado dentro de cada uno de los *sprints*, puesto que son los usuarios los que validan el resultado final. Las reuniones con los usuarios permiten establecer los productos para el nuevo *sprint*, en el cual se registran mejoras, errores o inconsistencias.

#### 4. CONCLUSIONES Y TRABAJOS FUTUROS

Es importante reconocer que un data warehouse es alimentado por todos los datos que resultan de la operación diaria de los sistemas. Todos estos datos deberán pasar por procesos de limpieza para que sean de utilidad para la toma de decisiones.

Para el caso de la evaluación del desempeño del docente en la Universidad de Cuenca, los usuarios han quedado satisfechos, debido a que luego de realizar una validación del producto, han concluido que ahora cuentan con los datos necesarios para poder analizar la información de cada proceso de evaluación de manera más rápida y eficaz.

La arquitectura definida para la construcción del data warehouse fue bastante útil, puesto que la aplicación de la metodología Hefesto combinada con la metodología de desarrollo Scrum, permitió una mejor planificación y avance del trabajo. La colaboración cercana y activa de los usuarios ayudó a corregir incidentes y errores sin pérdida de tiempo ni de recursos.

El uso de la notación BPMN para modelar los procesos ETL para la integración de datos, al ser una notación estándar y gráfica ayudó a que los usuarios puedan comprender de manera fácil y rápida cada uno de los procesos.

En cuanto a los trabajos futuros, se pretende refinar el proceso dentro de la arquitectura propuesta, como, por ejemplo, definir *sprints* en base a todos los pasos de Hefesto, incluyendo también el modelamiento del BPMN del proceso ETL. Esta unificación total permitirá definir un proceso genérico de creación de un data warehouse para universidades, además se pretende probar en otros dominios.

La arquitectura definida será utilizada para la implementación de todos los *datamarts* para las áreas analizadas en este trabajo.

#### AGRADECIMIENTOS

Este trabajo fue realizado como parte de los requisitos, previo a la obtención del título “Master en Gestión Estratégica de Tecnologías de la Información” impartido por la Facultad de Ingeniería de la Universidad de Cuenca.

#### REFERENCIAS

Bernabeu, R. D. (2010). *Hefesto. Data warehousing: Investigación y sistematización de conceptos. Hefesto: Metodología para la construcción de un data warehouse*. Córdoba, Argentina, 160 pp. Disponible en <https://www.businessintelligence.info/assets/hefesto-v2.pdf>

---

<sup>3</sup> <http://www.pentaho.com/v2>

- Castillo, I. (2014). *Diseño, elaboración y explotación de un data warehouse para una institución sanitaria*. Bachelor's thesis, Universitat Oberta de Catalunya, España, 114 pp. Disponible en <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/34101/8/icastillohTFC0614memoria.pdf>
- Guan, J., Nunez, W., & Welsh, J. F. (2002). Institutional strategy and information support: the role of data warehousing in higher education. *Campus-wide information systems*, 19(5), 168-174.
- Inmon, W. H. (2002). *Building the data warehousing*. Wiley Computer Publishing, 428 pp. Disponible en <http://fit.hcmute.edu.vn/Resources/Docs/SubDomain/fit/ThayTuan/DataWH/Bulding%20the%20Data%20Warehouse%204%20Edition.pdf>
- Lozada Peñafiel, X. N., & Cruz Tamayo, H. D. (2014). *Análisis, diseño, construcción e implementación de un data warehouse para toma de decisiones y construcción de los KPI, para la empresa Kronosconsulting Cia Ltda.* Disertación doctoral, Carrera de Ingeniería en Sistemas e Informática, Universidad de las Fuerzas Armadas ESPE, Quito, Ecuador.
- Menéndez, M., & de Lujan Gurmendi, M. (2012). *Sistemas para la toma de decisiones en el ámbito universitario*. 6° Simposio Argentino De Informática En El Estado, Buenos Aires, Argentina, 12 pp. Disponible en [http://41jaiio.sadio.org.ar/sites/default/files/333\\_SIE\\_2012.pdf](http://41jaiio.sadio.org.ar/sites/default/files/333_SIE_2012.pdf)
- Pirani, J. A., & Albrecht, B. (2005). *University of Phoenix: Driving decisions through academic analytics*. EDUCAUSE Center for Applied Research, Colorado, USA, 17 pp. Disponible en <https://net.educause.edu/ir/library/pdf/ers0508/cs/ECS0509.pdf>
- Ross, M., & Kimball, R. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3<sup>rd</sup> ed.). John Wiley & Sons Inc., 601 pp. Disponible en <http://www.essai.rnu.tn/Ebook/Informatique/The%20Data%20Warehouse%20Toolkit,%203rd%20Edition.pdf>