

HACIA UN SISTEMA DE AYUDA A LA DECISIÓN PARA UNIVERSIDADES: CASO DE USO DE LA UNIVERSIDAD DE CUENCA

Carmen Rojas Muñoz^a, Víctor Saquicela Galarza^b,

^a Universidad de Cuenca, Dirección de Tecnologías de Información y Comunicación,
Cuenca, Ecuador

carmita.rojas@ucuenca.edu.ec

^b Universidad de Cuenca, Departamento de Ciencias de la Educación, Cuenca, Ecuador

victor.saquicela@ucuenca.edu.ec

Resumen. Actualmente la Universidad de Cuenca cuenta con 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 lo que, emitir informes integrados a partir de los sistemas se ha convertido en todo un reto. Por lo tanto, en este trabajo se propone la creación de un DataWarehouse donde se integre los datos de todos los sistemas informáticos para poder emitir informes consolidados y a futuro poder extraer conocimiento. Para lograr esta integración se propone la utilización conjunta de metodologías y tecnologías de: creación de DataWarehouse, desarrollo ágil (SCRUM) y BPMN.

Palabras Clave: DataWarehouse, Datamart, Hefesto, decisiones, indicadores, educación superior

Eje temático: Este trabajo ha sido realizado con la finalidad de proveer ayuda a en la Gestión, por tanto, se posiciona en el eje temático: Soluciones TIC para la Gestión.

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, información que es necesaria para que las autoridades puedan tomar decisiones, para obtener la información correspondiente a 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, convirtiéndose en la herramienta ideal para la toma de decisiones en cualquier área de la Institución basado 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, a través del presente trabajo se planteó la elaboración de un DataWarehouse 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 DataWarehouse 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 ser replicado por otras por todas la Universidades del País.

El resto del trabajo está organizado de la siguiente manera: la sección 2 describe los antecedentes referentes a los conceptos de DataWarehouse, 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 del DataWarehouse 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 DataWarehouse y se analizan algunos trabajos relacionados que guiarán la descripción del trabajo.

2.1 Antecedentes

Para el desarrollo de un DataWarehouse, existen algunas metodologías, entre ellas están: Inmon, Kimball [1] y Hefesto [5]. La metodología de Bill Inmon señala que se debe partir de un DataWarehouse para luego llegar a los Datamarts [17], es decir será una metodología top-down; por otro lado, Ralph Kimball propone su metodología de forma contraria a Inmon es decir se partirá de los Datamarts para llegar a construir el DataWarehouse [4], siendo entonces una metodología bottom-up. Por otro lado, la metodología Hefesto, la cual se escogió para el desarrollo de este trabajo, por su parte: plantea una combinación entre las dos metodologías (Inmon y Kimball) y permite implementar el DataWarehouse partiendo de los requerimientos de los usuarios para construir un esquema lógico, y la definición de los procesos ETL [5]. 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 DataWarehouse. 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, a partir del modelo conceptual se define 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 DataWarehouse se almacenen los datos sin errores.

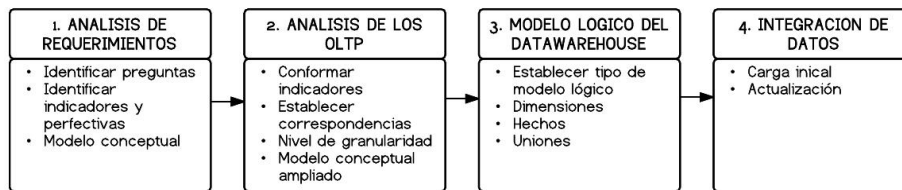


Figura 1: Metodología Hefesto

Al hablar del desarrollo de proyectos de software, existen varias metodologías, sin embargo, una metodología ágil que permite interactuar con el cliente y entregar resultados parciales de manera que éste se sienta satisfecho y comparta sus inquietudes, incluso para solicitar cambios durante el desarrollo del proyecto. SCRUM es una metodología ágil que 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 DataWarehouse (Hefesto) de manera ágil (SCRUM) que permita a futuro ser reproducido por otras universidades.

2.1 Trabajos relacionados

Se analizaron varios casos sobre el uso de un DataWarehouse en una empresa, concluyendo que este tipo de soluciones son útiles para la toma de decisiones. Se pudo observar que un DataWarehouse puede ser aplicado en cualquier empresa no importa cuál sea su actividad, se pudo ver que existen este tipo de sistemas en empresas de actividad financiera [6] para examinar a sus deudores y cartera vencida, o para empresas sanitarias [9], en donde se considera fundamental dotar de un sistema para la toma de decisiones para los directivos. 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 [10], este proyecto señala que trabajar con un DataWarehouse simplifica procesos para tomar decisiones, puesto que el acceso es fácil y es beneficios para el usuario que pueda hacer sus propias consultas. Se revisó también el análisis del rol del datawarehousing en la educación superior [14], en este artículo se examina las metas y retos de las universidades con respecto al manejo de la información, necesaria para la toma de decisiones y para planificar estratégicamente. Plantea que un DataWarehouse 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. Un paper más analizado de la Universidad de Phoenix: "Driving decisions through academic analytics [15], en este estudio se indica que todas las Universidades buscan tener una herramienta que les permita realizar el análisis de información relevante que le permita tomar decisiones acertadas. Luego de este trabajo pudieron observar 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, se pudo ver que un DataWarehouse hoy en día es de gran ayuda para la toma de decisiones de los más altos directivos de una empresa, por lo que se llegó a la conclusión de que el desarrollo e implementación del DataWarehouse 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 DataWarehouse

En esta sección se describe la arquitectura propuesta para el desarrollo del DataWarehouse, para ello se utiliza un ejemplo práctico sobre un Datamart específico. Este Datamart guiará todo el proceso propuesto.

3.1 Arquitectura del DataWarehouse

La figura 3 visualiza la arquitectura definida para la construcción del DataWarehouse 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 DW. Esto se hizo con el objetivo de obtener un producto que esté acorde a las necesidades del usuario, puesto que con la participación activa 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 como se modeló cada uno de los procesos ETL para la construcción del DataWarehouse. Con la combinación descrita anteriormente, se pretende demostrar que el uso de SCRUM conjuntamente 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 DataWarehouse.

A continuación, se explica cada uno de los componentes de la arquitectura; aplicados en la construcción del Datamart que contiene los datos de la evaluación del desempeño del docente. Todos los Datamarts definidos fueron construidos aplicando esta metodología.

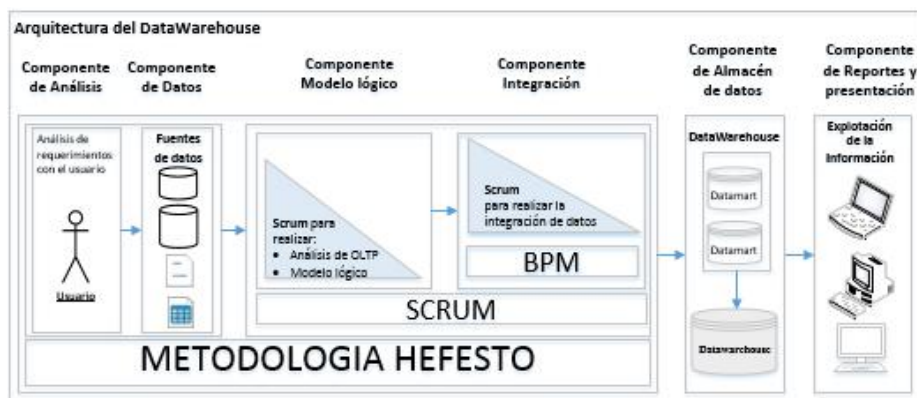


Figura 3: Arquitectura del DataWarehouse

3.1.1 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. Resultado de este proceso se obtuvo 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)¹ y el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEACES)² 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 DataWarehouse

Dependencias	DataMart	Indicadores
<ul style="list-style-type: none"> Facultades 	Datamart de Grado o carreras	Promedios de estudiantes por periodo lectivo y por carrera

¹ <http://www.ces.gob.ec>

² <http://www.ceaaces.gob.ec>

<ul style="list-style-type: none"> • Consejo Académico • Direcciones de carrera 		Número de estudiantes por grupo ofertado y por docente en un periodo lectivo
		Número de Matriculados x Carrera en un periodo lectivo
		Estudiantes nuevos matriculados por carrera en un periodo lectivo
		Número de estudiantes graduados por periodo lectivo y por carrera
<ul style="list-style-type: none"> • Comisión de Evaluación Interna 	Datamart de Evaluación del docente	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
<ul style="list-style-type: none"> • Dirección de Talento Humano 	Datamart 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
		Número de servidores universitarios con capacidades diferentes
<ul style="list-style-type: none"> • Dirección de Posgrados 	Datamart 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
<ul style="list-style-type: none"> • Dirección de Evaluación • Dirección de Planificación 	Datamart de Proyectos	Número de participantes por proyecto por dependencia y por año
		Total de presupuesto por proyecto por dependencia y por año
<ul style="list-style-type: none"> • Dirección de Investigación 	Datamart 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
<ul style="list-style-type: none"> • Calificación • Porcentaje de participación 	<ul style="list-style-type: none"> • Docentes • Facultades • Carreras • Actores • Funciones • Tiempo (proceso de evaluación)

3.1.2 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.

3.1.3 Componente modelo lógico

Continuando con los pasos de la metodología Hefesto (segundo paso), lo siguiente es la construcción del modelo conceptual del Datamart a partir de los indicadores y perspectivas identificadas en el paso anterior. Para el componente del modelo lógico se inició con el uso de la metodología de desarrollo de proyectos SCRUM, con una iteración para el análisis de los OLTP y la elaboración del modelo lógico. Es decir, basado en la tabla 1 se crea un listado de productos (análisis de OLTP, elaboración del modelo lógico, granularidad, etc) que se insertan en un sprint, el cual será ejecutado por un grupo de personas (5).

Continuando con la aplicación de la metodología para el caso de estudio, en la figura 4 se puede visualizar el modelo lógico, en donde se indica que la relación que une las perspectivas con los indicadores requeridos por los usuarios se la llama evaluación.

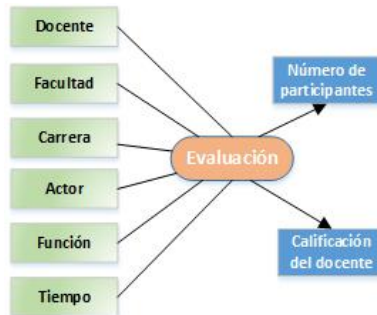


Figura 4: Modelo conceptual del Datamart de evaluación

Otra actividad que es parte del componente modelo lógico y que se realizó durante el primer sprint fue el análisis de las fuentes OLTP con el fin de determinar la relación entre el modelo lógico y las fuentes de datos; y para determinar cómo se conformaron cada uno de los indicadores identificados por los usuarios. En la tabla 3, se puede observar cómo se realizó la conformación de los indicadores para el caso del Datamart de evaluación.

- “Número de participantes”
Este indicador se obtiene de contar cuantos evaluadores y evaluados participaron 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	$\frac{(D \times H.D) + (I \times H.I) + (D.G. \times H.D.G)}{\text{Dedicación docente}}$
<p><i>D = Docencia</i> <i>H.D = Total de horas de docencia asignadas al docente en su distributivo</i> <i>I = Investigación</i> <i>H.I = Total de horas de investigación asignadas al docente en su distributivo</i> <i>D.G = Gestión académica</i> <i>H.D.G = Total de horas de gestión y dirección asignadas al docente en su distributivo</i> 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</p>	

Una vez que se estableció el modelo lógico y se realizó el análisis de los datos en el primer componente se procedió a establecer las correspondencias entre las fuentes de datos y el modelo conceptual, tal como se puede observar en la figura 5.

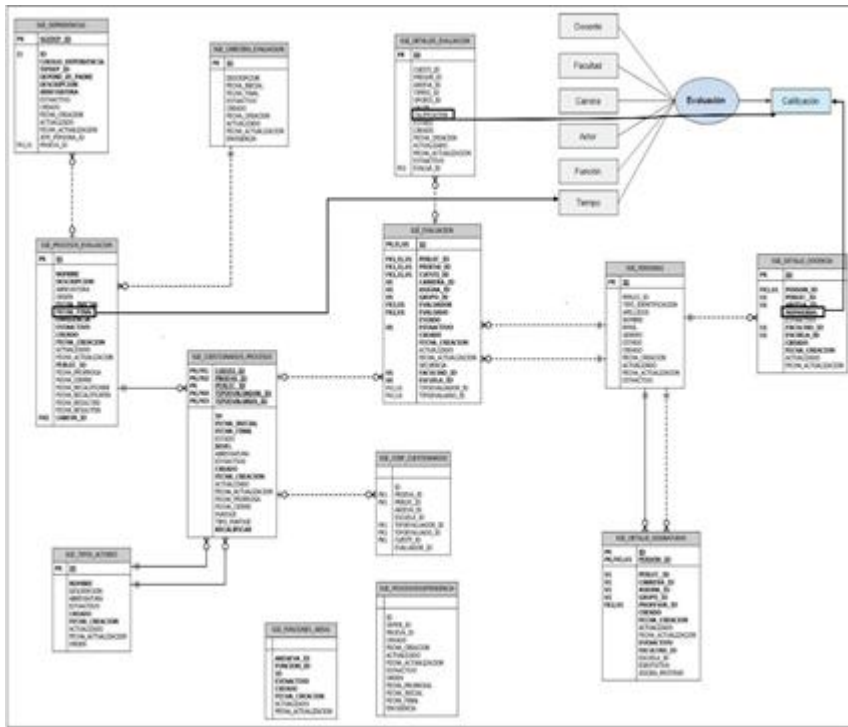


Figura 5: Correspondencias para el indicador de calificaciones del Datamart de evaluación

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 como 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 para las perspectivas docente y facultad para el Datamart de evaluación

Perspectiva	Campos	Diagrama entidad-relación
Docente	<ul style="list-style-type: none"> • person_id • apellidos • nombres • edad 	adminuc.personas_naturales
Facultad	<ul style="list-style-type: none"> • descripción 	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.

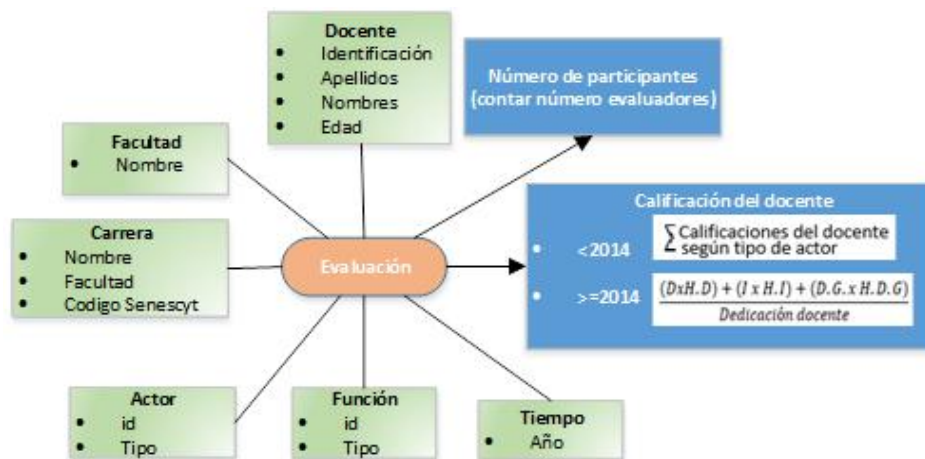


Figura 6. Modelo conceptual ampliado para el Datamart de evaluación

Para finalizar el sprint, la siguiente actividad fue confeccionar el modelo lógico de la estructura del DataWarehouse, 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.

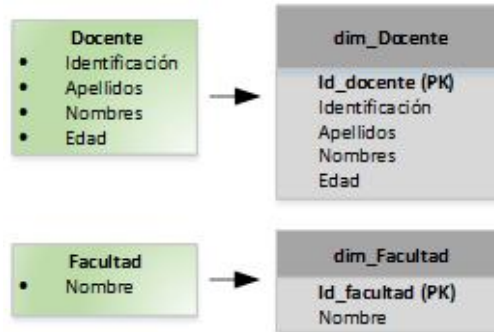


Figura 7. Ejemplos de tablas de dimensiones para el Datamart de evaluación

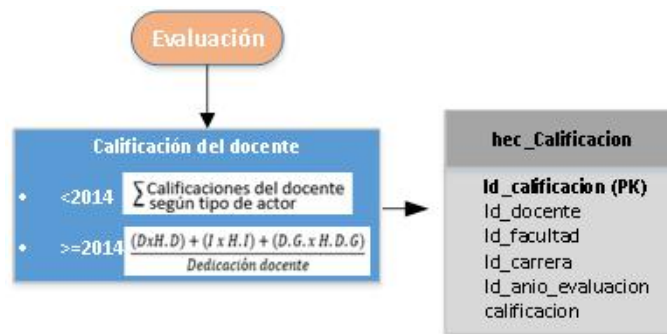


Figura 8. Tabla de hechos para obtener la calificación del docente para el Datamart de evaluación

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.

3.1.4 Componente Integración

Una vez que se tiene el modelo lógico, en el componente de integración se procede a poblar el DataWarehouse con los datos obtenidos de los sistemas operacionales, sin embargo, los datos deben pasar por un proceso (ETL) de limpieza para evitar datos anómalos que no deben existir en el DW. En este componente se continúa con la metodología Hefesto y con el uso de SCRUM, no obstante, en este componente 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 con la notación BPMN el proceso ETL para la carga de datos de las calificaciones de los docentes y la figura 10 muestra el proceso ETL que se ejecutó para la carga inicial. Una vez pobladas las dimensiones y hechos, se procedió a la creación de los cubos multidimensionales. La figura 11, muestra un ejemplo del cubo llamado calificaciones.

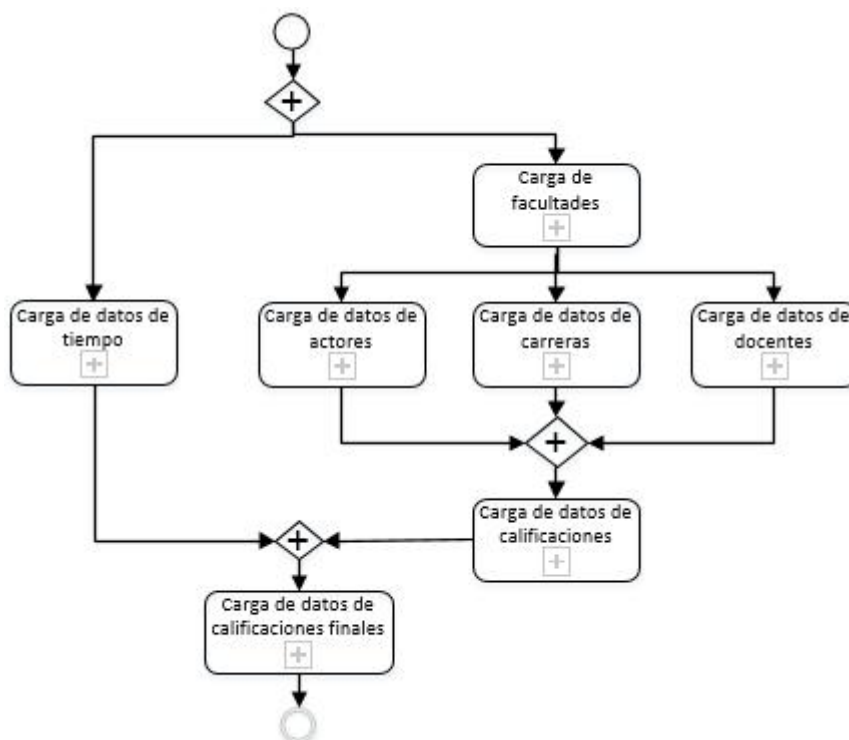


Figura 9. Modelado con BPMN del proceso ETL para calificaciones

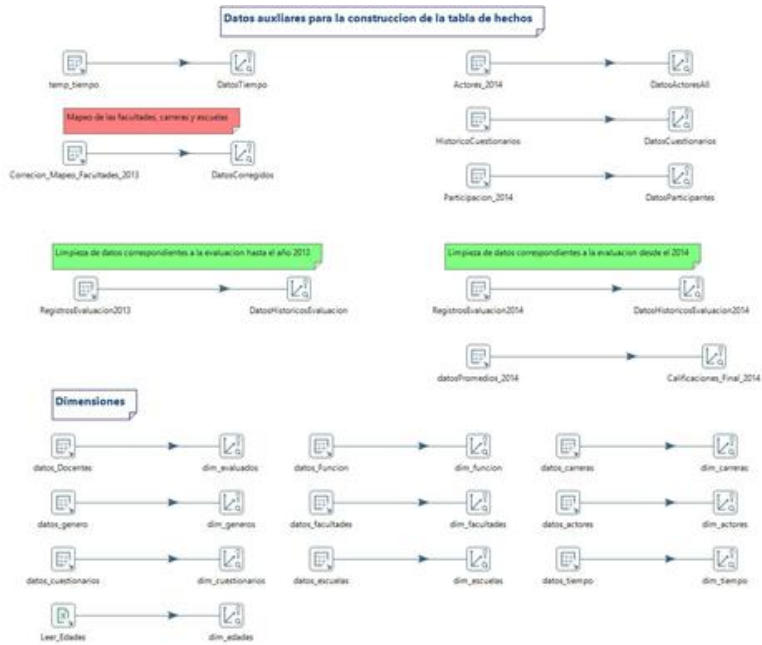


Figura 10. Elaboración del proceso ETL para la carga de datos

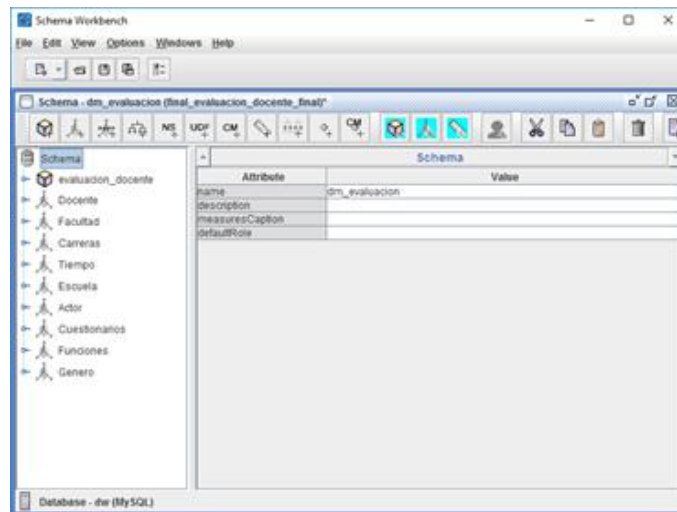


Figura 11. Cubo elaborado para las calificaciones de los docentes

Los procesos ETL modelados con BPMN se han ejecutado 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.

3.1.5 Componente de almacén de datos

El componente de almacén de datos, se refiere al DataWarehouse 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 DataWarehouse. Para el caso del ejemplo durante todo este proyecto este componente se trata del datamart de evaluación. En el caso específico de la Universidad de Cuenca, el almacenamiento centralizado de datos se realiza utilizando un modelo de DataWarehouse 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.

3.1.6 Componente de Reportes y presentación

Este componente es el más importante para los usuarios, puesto que es a través de éste donde 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 Pentaho³, 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 para 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 por cada uno de los datamarts implementados.

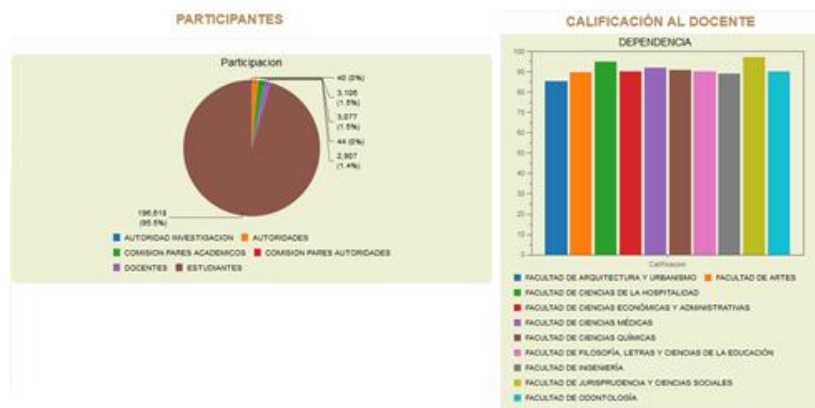


Figura 12. Ejemplos de reportes que se encuentran para el Datamart de evaluación

³ <http://www.pentaho.com/v2>

Este componente también está considerado dentro de cada uno de los sprints, puesto que son los usuarios los que validan el resultado final. Producto de reuniones con los usuarios permite 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 DataWarehouse es alimentado por todos los datos que resultan de la operación diaria de los sistemas; sin embargo, 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 DataWarehouse 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, con la colaboración cercana y activa de los usuarios, ayudando de esta manera a corregir incidentes y errores a tiempo, 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 los sprints basado en 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 DataWarehouse para universidades, además se pretende probar en otros dominios.

5 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, por esta razón a través del presente se expresa un profundo agradecimiento a todos quienes son parte de la Dirección de posgrados de la facultad de Ingeniería y a la Facultad como tal.

6 Referencias

1. Breslin, M. (2004). Data warehousing battle of the giants. *Business Intelligence Journal*, 7.
2. Calabria Sarmiento, J. C. (2011). Construcción y poblamiento de un DataWarehouse basado en el paradigma de bases de datos objeto relacional.
3. Brizuela, E. I. L., & Blanco, Y. C. (2013). Metodologías para desarrollar Almacén de Datos. *Revista Arquitectura e Ingeniería*, 7(3), 3-12.
4. Ross, M., & Kimball, R. (2013). *The data warehouse toolkit: the complete guide to dimensional modeling*. Wiley.
5. Bernabeu, R. D. (2010). Hefesto–Data warehousing: Investigación y sistematización de conceptos-Hefesto: Metodología para la construcción de un data warehouse. Córdoba, Argentina.
6. 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 (Doctoral dissertation, Universidad de las Fuerzas Armadas ESPE. Carrera de Ingeniería en Sistemas e Informática.).
7. Aránega Hernández, A. (2015). Almacenes de datos: Análisis de ventas de una compañía (Master's thesis, Universitat Oberta de Catalunya).
8. Castillo Hernández, I. (2014). Diseño, elaboración y explotación de un data warehouse para una institución sanitaria (Bachelor's thesis, Universitat Oberta de Catalunya).
9. Menéndez, M., & de Lujan Gurmendi, M. (2012). Sistemas para la toma de decisiones en el ámbito universitario. Coordinador Sistemas para la toma de decisiones–Consortio SIU, Directora Ejecutiva–Consortio SIU.
10. La Red Martínez, D. L., Acosta, J. C., Cutro, L. A., Uribe, V. E., & Rambo, A. R. (2010). Data warehouse y data mining aplicados al estudio del rendimiento académico y de perfiles de alumnos. In XII Workshop de Investigadores en Ciencias de la Computación.
11. Matamala, C. Z., Díaz, D. R., Cuello, K. C., & Leiva, G. A. (2011). Análisis de rendimiento académico estudiantil usando data warehouse y redes neuronales/Analysis of students' academic performance using data warehouse and neural networks. *Ingeniare: Revista Chilena de Ingeniería*, 19(3), 369.
12. Dell'Aquila, C., Di Tria, F., Lefons, E., & Tangorra, F. (2007, August). An academic data warehouse. In *Proceedings of the 7th WSEAS International Conference on Applied Informatics And Communications* (pp. 24-26).
13. 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.
14. Pirani, J. A., & Albrecht, B. (2005). University of Phoenix: Driving decisions through academic analytics. EDUCAUSE Center for Applied Research.
15. Bustamante, J. T. (2015). Inteligencia de Negocios. Una herramienta de apoyo para la toma de decisiones en operaciones militares.
16. Bissi, W. (2007). SCRUM-Metodología de desenvolvimiento ágil. *Campo Digital*, 2(1), 03-06.
17. Inmon, W. H. (2002). *Building the Data Warehousing*.
18. Gallego, M. T. (2012). Metodología Scrum. *Gestión de Proyectos Informáticos*, <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/17885/1/mtrigasTFC0612memoria.pdf>