

UNIVERSIDAD DE CUENCA



FACULTAD DE INGENIERÍA CENTRO DE POSGRADOS

MAESTRÍA EN GESTIÓN ESTRATÉGICA DE TECNOLOGÍAS DE LA INFORMACIÓN

DATAWAREHOUSE para la Universidad de Cuenca: Indicadores para
la toma de decisiones

**TRABAJO DE GRADUACIÓN PREVIO A LA
OBTENCIÓN DEL GRADO DE MAGISTER
EN GESTIÓN ESTRATÉGICA DE
TECNOLOGÍAS DE LA INFORMACIÓN**

Autor: Ing. Carmen Rojas Muñoz. C.I.: 0103691671

Director: Ing. Víctor Saquicela Galarza, PhD. C.I.: 0103599577

CUENCA – ECUADOR

Junio – 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 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



ABSTRACT

Currently the University of Cuenca has several information systems that generate large amounts of data, this represents a heterogeneity from different points of view such as: database, programming languages, schemas, data, etc. This various sources causes that each computer system will operate as stand – alone system’, when all the system access to the data simultaneously causes a bottleneck, because each system has its own list of reports. This makes a challenge to give a report from the integrated systems. Therefore, in this paper is proposed the setting-up of a DataWarehouse, with this all the data from the entire computer systems are integrated to be able to supply complete reports and also function as historic database. To achieve this integration, it is proposed to unify methodologies and technologies’ usage to create DataWarehouse, Agile Development (SCRUM) and BPMN

Keywords: DataWarehouse, Datamart, Hefesto, Decisions, Indicators, Higher education



DEDICATORIA

A mi esposo por su apoyo incondicional, a mi hija por ser mi inspiración, a mi madre por su dedicación y amor, a mi padre porque a pesar de que ya no está con nosotros físicamente, ha dejado su legado de lucha y valentía en mí y siempre estará en mi corazón, a mis hermanos por estar siempre a mi lado, de manera especial a Remigio y Aida por todo el cariño y apoyo que me han brindado siempre, y a toda mi familia por haberme impulsado a cumplir mis metas.



AGRADECIMIENTO

Agradezco sobre todo a Dios por ser el guía y el pilar fundamental de mi vida en todo momento. Agradezco luego a mi esposo y a mi hija por toda la paciencia y amor que me ha brindado para lograr esta meta. Agradezco también a toda mi familia y a mi suegra por el constante aliento y apoyo que me brindaron durante todo el tiempo que duró mis estudios. Un agradecimiento especial al Ing. Víctor Saquicela quien ha confiado en mí, y con su capacidad y experiencia ha sabido guiarme para expresar y plasmar mis ideas en este trabajo, siendo un aporte invaluable en la culminación de esta tarea. Por último, agradezco a todas aquellas personas que de una u otra manera me han apoyado y colaborado para llegar al fin de este proyecto.



Contenido

Listado de Figuras.....	8
Listado de Tablas	9
Cláusulas de derecho de autor	10
Cláusulas de propiedad intelectual.....	11
Capítulo 1. INTRODUCCIÓN.....	12
1.1. Antecedentes	13
1.2. Definición del problema.....	13
1.3. Justificación.....	14
1.4. Objetivos	14
1.4.1. General	14
1.4.2. Específicos	15
1.5. Alcance.....	15
Capítulo 2. MARCO TEÓRICO Y TRABAJOS RELACIONADOS.....	17
2.1. Sistemas de información para la toma de decisiones	17
2.2. Inteligencia de negocios	18
2.3. OLTP vs OLAP	18
2.4. DataWarehouse	19
2.4.1. Definición de DataWarehouse.....	19
2.4.2. Definición de Datamart	20
2.4.3. Modelo Multidimensional	20
2.4.4. Metodología Hefesto	23
2.4.5. Explotación de la información	28
2.5. Metodologías de desarrollo de proyectos de software	30
2.5.1. Metodologías tradicionales.....	30
2.5.2. Metodologías ágiles.....	31
2.6. Modelamiento de procesos de negocio utilizando notación BPMN	32
2.7. Trabajos relacionados.....	33
Capítulo 3. CONSTRUCCIÓN E IMPLEMENTACIÓN DEL DATAWAREHOUSE	36
3.1. Arquitectura del DataWarehouse	36
3.1.1. Componente de análisis.....	38
3.1.2. Componente de datos	42
3.1.3. Componente Modelo Lógico.....	43
3.1.4. Componente Integración de datos	49
3.1.5. Componente almacén de datos	54



3.1.6.	Componente de validación de resultados	56
3.1.7.	Componente de reportes y presentación.....	56
3.2.	Aplicación de la metodología SCRUM.....	57
Capítulo 4. VALIDACIÓN DE RESULTADOS		60
4.1	Verificación a través de las preguntas planteadas	60
4.1.1	Primera pregunta: “Calificación obtenida por un docente en una facultad y carrera en un tiempo determinado”	60
4.1.2	Segunda pregunta: “Calificación obtenida por el docente de parte de un actor en una función en un tiempo determinado”	61
4.1.3	Tercera pregunta: “Número de participantes por facultad y género en un tiempo determinado”	62
4.2	Pruebas de calidad de la solución.....	63
4.2.1	Pruebas unitarias y de integración.....	63
4.2.2	Pruebas de validación de datos.....	64
4.2.3	Pruebas de rendimiento	64
4.2.4	Pruebas de regresión.....	65
4.2.5	Pruebas de aceptación de usuario.....	65
Capítulo 5. CONCLUSIONES Y TRABAJOS FUTUROS.....		68
Glosario de términos.....		70
Referencias bibliográficas.....		71
ANEXO 1. Identificación de la Empresa - UNIVERSIDAD DE CUENCA		73
Misión		73
Visión		73
Políticas Institucionales.....		73
Valores		74
Estrategias y líneas de acción.....		75
Entorno.....		77
Estructura Organizacional de la Universidad de Cuenca		77
Relación de los objetivos estratégicos con los objetivos del DataWarehouse		79
Áreas de aplicación del DataWarehouse		79
ANEXO 2. Procesos ETL.....		80
Modelamiento BPMN		80
Carga de la dimensión: dim_Tiempo		80
Carga de la dimensión: dim_Docente		80
Carga de las dimensiones: dim_Facultad y dim_Escuela		81
Carga de las dimension: dim_Actor		81



Carga de las dimension: dim_Edad	81
Carga de las dimension: dim_Genero	82
Limpieza y carga de las evaluaciones docentes	82
Procesos ETL en herramienta Pentaho	83
ETL para las Dimensiones	83
ETL para Tabla de hechos.....	83
ANEXO 3. Base de datos para el Sistema de Gestión de Evaluación del docente	84
Diccionario de Datos.....	85



Listado de Figuras

Figura 1: Piramide Organizacional.	17
Figura 2: OLTP vs OLAP	19
Figura 3: Modelo Estrella.....	22
Figura 4: Modelo copo de nieve.....	23
Figura 5: Pasos de la metodología Hefesto.	24
Figura 6: Ejemplo de modelo conceptual.....	25
Figura 7: Ejemplo de modelo conceptual ampliado.....	26
Figura 8: Metodología ágil de desarrollo de proyectos SCRUM.....	32
Figura 9: Ejemplos de elementos de la Notación de BPMN.....	33
Figura 10: Arquitectura del DataWarehouse.....	37
Figura 11: Indicadores y perspectivas correspondientes al datamart de evaluación	41
Figura 12: Modelo conceptual del Datamart de evaluación.....	42
Figura 13: Fuentes de datos para el Datamart de evaluación	43
Figura 14: Correspondencias para el Datamart de evaluación	45
Figura 15: Modelo conceptual ampliado para el Datamart de evaluación	47
Figura 16: Tablas de dimensiones para el Datamart de evaluación	48
Figura 17: Tablas de hechos para el Datamart de evaluación	48
Figura 18: Uniones del Datamart de evaluación	49
Figura 19: Vista general del proceso de carga de datos de evaluación	51
Figura 20: Procesos ETL para la limpieza de datos	52
Figura 21: Ejemplo de un cubo multidimensional	54
Figura 22: DataWarehouse de la Universidad de Cuenca.....	55
Figura 23: Ejemplos de reportes obtenidos a partir del Datamart de evaluación	57
Figura 24: Organigrama de la Universidad de Cuenca	78
Figura 25: Carga para la dimensión tiempo (procesos de evaluación).....	80
Figura 26: Carga para la dimensión docente	80
Figura 27: Carga para las dimensiones facultad y escuela.....	81
Figura 28: Carga para la dimensión actor	81
Figura 29: Carga para la dimensión edad.....	81
Figura 30: Carga para la dimensión genero	82
Figura 31: Limpieza y carga de datos para las evaluaciones	82
Figura 32: ETL para dimensiones	83
Figura 33: ETL para tabla de hechos	83
Figura 34: Modelo Entidad – Relación SGE.....	84



Listado de Tablas

Tabla 1: Datamarts e indicadores establecidos para el DataWarehouse	39
Tabla 2: Indicadores y perspectivas para el Datamart de evaluación.....	41
Tabla 3: Actividades planificadas para el sprint del componente modelo lógico	44
Tabla 4: Conformación de indicadores para el Datamart de evaluación	45
Tabla 5: Nivel de granularidad para las perspectivas del Datamart de evaluación	46
Tabla 6: Actividades planificadas para el sprint del componente de integración de datos	50
Tabla 7: Actividades ejecutadas en el primer sprint.....	58
Tabla 8: Actividades ejecutadas en cada sprint del componente de integración.....	58
Tabla 9: Calificación del docente por facultad, carrera y proceso de evaluación	61
Tabla 10: Calificación de un docente por actor y función en un proceso de evaluación	62
Tabla 11: Número de participantes por facultad y género en un proceso de evaluación	63
Tabla 12: Cuestionario aplicado a usuario final	65
Tabla 13: Contabilización de respuestas al cuestionario	66
Tabla 14: Resumen de las respuestas de manera clasificada.....	66
Tabla 15: Diccionario de datos del esquema entidad-relacion del SGE.....	85



Cláusulas de derecho de autor



Universidad de Cuenca
Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Carmen del Rocío Rojas Muñoz en calidad de autora y titular de los derechos morales y patrimoniales del trabajo de titulación "DATAWAREHOUSE para la Universidad de Cuenca: Indicadores para la toma de decisiones", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 10 de junio de 2017





Cláusulas de propiedad intelectual



Universidad de Cuenca
Cláusula de propiedad intelectual

Carmen del Rocío Rojas Muñoz, autora del Trabajo de Titulación "DATAWAREHOUSE para la Universidad de Cuenca: Indicadores para la toma de decisiones", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autora.

Cuenca, 10 de junio de 2017

Carmen del Rocío Rojas Muñoz

C.I: 0103691671



Capítulo 1. INTRODUCCIÓN

Hoy en día, la Universidad de Cuenca cuenta con 16 sistemas informáticos que generan una 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. 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 basadas en los datos. Además, permite obtener los indicadores establecidos por los organismos de gobierno y que faciliten la realización de 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 analizar la información desde diferentes puntos de vista, convirtiéndose en la herramienta ideal para la toma de decisiones en cualquier área de la Institución basada en información integrada y global. A través del presente trabajo se planteó la elaboración de un DataWarehouse para la Universidad de Cuenca, el mismo que ha sido 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 proceso para la construcción de un sistema de ayuda a la toma de decisiones para la Universidad de Cuenca y que a futuro puede ser implementado en otras universidades del país.

Al finalizar este trabajo, la Universidad de Cuenca obtuvo un sistema para la toma de decisiones que fue construido a través de la implementación de un proceso combinando las metodologías: Hefesto para la construcción de un DataWarehouse, SCRUM para el desarrollo de proyectos de software y la notación BPMN para modelar los procesos.



1.1. Antecedentes

Según el Programa de las Naciones Unidas para el Desarrollo (PNUD): “La educación es el medio principal para la construcción de valores para la sociedad”. El gobierno ecuatoriano, apegado a esta afirmación, habla sobre el “Principio de calidad” y la “Evaluación de la calidad” en los artículos 93 y 94 de la Ley de Orgánica de Educación Superior (Ley Orgánica de Educación Superior, LOES, 2010). Por esta razón, cada una de las Universidades del país son sometidas a procesos de evaluación y como resultado de estos procesos son clasificadas y categorizadas. En el año 2008 se realizó en el país una primera clasificación de las universidades ecuatorianas y como resultado: siete universidades se clasificaron en la clase “A”, entre ellas la Universidad de Cuenca, sin embargo, en una segunda evaluación realizada en el año 2013, la Universidad de Cuenca pierde su categoría para ubicarse en la categoría “B”, debido a que no superó el análisis según los criterios de academia, eficiencia académica, investigación, organización e infraestructura.

El plan estratégico 2012-2017 de la Universidad contempla que en el año 2017 será una Universidad de docencia con investigación y se ha planteado como objetivo llegar a ser una Institución de Educación Superior de excelencia a nivel nacional e internacional, para ello debe cumplir los diferentes aspectos considerados en el marco de referencia establecido por el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES). Por lo antes mencionado, es necesario que la Universidad de Cuenca implemente una herramienta para integrar, depurar y procesar los datos, de manera que permita 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, basándose en información integrada y global, de manera que se pueda realizar siempre una autoevaluación para identificar problemas e inconvenientes y tratar de corregirlos y por tanto permitir una mejora continua en todos los procesos de la Universidad de Cuenca.

1.2. Definición del problema

Al igual que en toda empresa, en la Universidad de Cuenca, la información es uno de los activos intangible más importantes, es por esto, que su correcto manejo es de vital importancia para la toma de decisiones. Uno de los objetivos de la Universidad de Cuenca es ser una Universidad de excelencia tanto a nivel nacional como internacional, para ello las autoridades deben tomar decisiones adecuadas y en el momento preciso, basadas en los datos existentes en los sistemas informáticos.



Hoy en día la Universidad no cuenta con sistemas totalmente integrados, razón por la cual no existe una base de datos centralizada, y por ende en ciertos momentos la información se debe obtener de manera manual, esto representa un riesgo el momento de tomar decisiones, debido a que la información no está disponible y/o confiable en el momento adecuado. Otro problema latente es que, el momento que las autoridades solicitan información, existe la dependencia de los técnicos que conocen el proceso y saben de manera precisa como generar la información necesitada, por lo que si el técnico encargado, no está en la Institución se produce un cuello de botella muy significativo e incluso retraso en los procesos.

1.3. Justificación

A través del presente trabajo se plantea la elaboración de un DataWarehouse el mismo que será implementado en base a un profundo análisis de las necesidades de la Institución, y que integrará toda la información y estará disponible en línea para todas las autoridades y quienes trabajen diariamente con la información para la toma de decisiones. Para que el trabajo sea un éxito, se contará con la participación activa de los usuarios de las áreas en los que es de suma urgencia contar con la información en tiempo real. Para la implementación de esta herramienta, se seguirá una metodología ágil de desarrollo para un proyecto de inteligencia de negocios, que permitirá analizar las fuentes y limpiar los datos para evitar inconsistencias, duplicidad, datos incompletos etc., para luego consolidar los datos en el DataWarehouse, creando los datamarts necesarios, que serán identificados durante el análisis preliminar del proyecto. A través de esta metodología se podrá descubrir posibles errores a tiempo sin necesidad de llegar al término del proyecto para que el usuario indique sus inconformidades por descubrir errores en la etapa final del proyecto.

1.4. Objetivos

1.4.1. General

Implementar un DataWarehouse para la Universidad de Cuenca que permita obtener de manera rápida y fiable la información, convirtiéndose en una herramienta de ayuda a las autoridades para la toma de decisiones. Esta información estará disponible para realizar una autoevaluación permanente de la



Institución, tomar decisiones y proyectar nuevos procesos que permitan siempre mejorar la Calidad de la educación en la Universidad. Además, el DataWarehouse permite obtener los indicadores propuestos por los organismos de gobierno, según los diferentes criterios como son: posgrado, dedicación de los docentes, evaluación del desempeño, resultados de investigación, eficiencia académica, entre otros.

1.4.2. Específicos

A continuación, se mencionan los objetivos específicos de este trabajo:

- Construir el DataWarehouse aplicando una metodología que sea fácil de implementar.
- Implementar el DataWarehouse como una herramienta de apoyo a la toma de decisiones.
- Contar con información analítica necesaria para obtener los indicadores correspondientes a los criterios y subcriterios de evaluación de la Universidad, mismos que podrán ser analizados por las distintas autoridades de la Institución.
- Agilizar las consultas y reportes que no se obtienen de los sistemas transaccionales de la Universidad.
- Proveer un repositorio en donde la información se mantenga disponible y confiable, puesto que se realizará un proceso de extracción, transformación y limpieza.
- Proveer una herramienta que permita reducir tiempos y costes en el proceso de extracción de la información, evitando procesos manuales y dependencias en los técnicos de la Institución.
- Aportar con el DataWarehouse al cumplimiento de los objetivos estratégicos planteados en el Plan Estratégico de la Universidad de Cuenca

1.5. Alcance

Es importante analizar la información financiera y académica de la Institución, con el objetivo de obtener los datos para cubrir todos los indicadores propuestos para la evaluación y posterior acreditación de la Institución y sus carreras, por este motivo el DataWarehouse almacenará la información sobre las siguientes áreas, cada una de las cuales conformará un datamart del DataWarehouse.:



- Sistema Académico de Grado
- Sistema Académico de Programas de Posgrados
- Sistema de Evaluación del Desempeño del Docente
- Sistema de Investigación
- Sistema de Talento Humano.
- Sistema de Proyectos

A través de este proyecto se implementarán los datamarts indicados, sin embargo, para la descripción de esta tesis se profundizará en el datamart correspondiente a la evaluación del desempeño del docente, con el objetivo de obtener los indicadores necesarios para que la dirección de evaluación analice todos los resultados de un proceso de evaluación y pueda dar a conocer los resultados a toda la comunidad universitaria. Se plasmará el proceso que se utilizará para definir las tareas de extracción, transformación y limpieza y de los datos para la creación de: los cubos OLAP, informes, y cuadros de mando para cada una de las áreas que conformarán el DataWarehouse de la Universidad de Cuenca.

Finalmente, una vez que la información haya sido extraída y consolidada en el datamart de evaluación, se procederá a realizar una validación de los resultados obtenidos con los usuarios, con el objetivo de constatar la validez del proyecto y sus resultados.

Capítulo 2. MARCO TEÓRICO Y TRABAJOS RELACIONADOS

En este capítulo se realizará una breve revisión de todos los conceptos necesarios para un mejor entendimiento de este trabajo.

2.1. Sistemas de información para la toma de decisiones

El objetivo principal de todo sistema de información es recolectar, procesar y almacenar información con el objetivo de optimizar los procesos de una empresa; sin embargo, un usuario que esté en los niveles más altos según la jerarquía de la empresa, necesita un sistema para la toma de decisiones. En la Figura 1, se puede observar una pirámide organizacional y cómo actúan los sistemas para los diferentes tipos de usuarios.

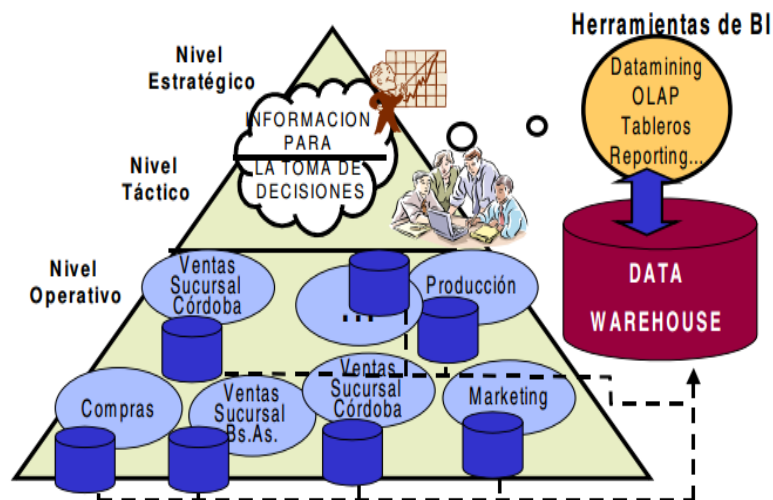


Figura 1: Pirámide Organizacional.

Fuente: Castellano, R. (2000). *La organización y el Data Warehouse*. [Figura]. Recuperado de <https://ti-1.wikispaces.com/5.1+Sistemas+de+Negocios>.

Como se puede observar para el nivel estratégico y nivel táctico de la empresa se necesitan los sistemas que ayudan a la toma de decisiones, y para el nivel operativo se utilizan los sistemas



transaccionales día a día. Por lo tanto, se puede concluir que un sistema para la toma de decisiones sirve de apoyo para que las autoridades de una empresa puedan tomar decisiones en base a un análisis de los datos provistos por estos tipos de sistemas.

2.2. Inteligencia de negocios

Los sistemas transaccionales operados por los usuarios de las empresas, generan grandes cantidades de datos, por lo que existe el reto de que todos estos datos se conviertan en información para que, a través de esta, los directivos de la empresa tengan el conocimiento necesario para que puedan planificar o tomar decisiones.

Según Curto (2012) “La inteligencia de negocios es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización” p.18. De esta definición se puede concluir que la inteligencia de negocios en una empresa ayuda a generar el conocimiento a través de la información resultante de los datos obtenidos desde los sistemas transaccionales, sin embargo, es necesario conocer cómo llevar los datos al conocimiento. Dos de los componentes de la inteligencia de negocios son: *DataWarehouse* el cual es el componente en donde estarán almacenados los datos y la *Minería de Datos* que será un método para analizar los datos del *DataWarehouse* (Sanchez, 2003). Los procesos de minería de datos están directamente relacionados al *DataWarehouse*, puesto que estos permiten un procesamiento más fácil y completo de la información (Inmon, *The data warehouse and data mining*, 1996).

2.3. OLTP vs OLAP

Los sistemas de información pueden ser divididos en dos clases: OLTP (**O**n **L**ine **T**ransaction **P**rocessing), estos son los sistemas transaccionales que se operan día a día en una empresa, cuyo fin es ayudar en la gestión de tareas propias de la empresa. La base de datos de este tipo de aplicaciones está diseñada para almacenar transacciones (Insert, Update, Delete) y no para analizar datos. Por otro lado, un Sistema OLAP (**O**n **L**ine **A**nalytical **P**rocessing), permite analizar los datos obtenidos a través de los OLTP, puesto que estos sistemas permiten extraer el conocimiento de información

histórica de la empresa para análisis, mejoras y para diseñar estrategias (Conn, 2005). En la Figura 2 se puede observar gráficamente la diferencia entre los sistemas OLTP y OLAP

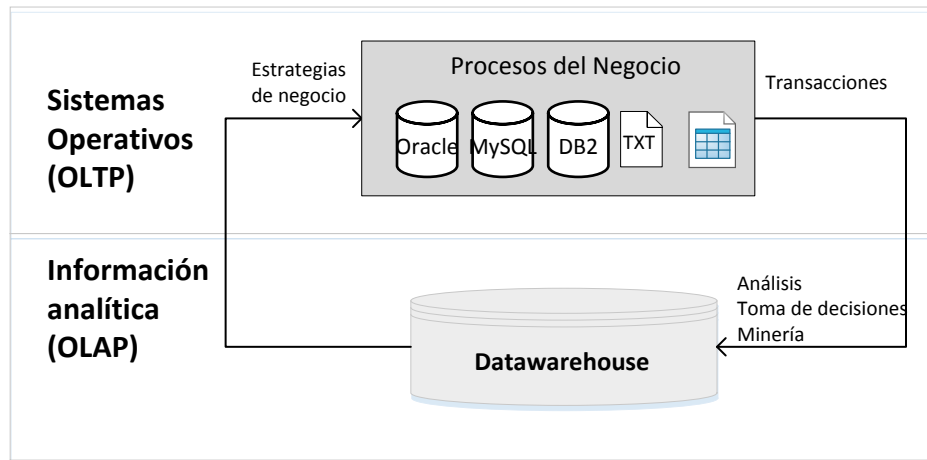


Figura 2: OLTP vs OLAP

Fuente: Elaboración propia basada en OLTP vs OLAP [Figura]. (2010) Recuperado de <http://datawarehouse4u.info/OLTP-vs-OLAP.html>

2.4. DataWarehouse

Hoy en día uno de los activos más importantes de una empresa es la información, siendo esta indispensable y necesaria para realizar análisis de la situación en la que se encuentra la empresa, con el objetivo de buscar estrategias para mejorar sus procesos y ser más competitiva. Para lograr este objetivo se debe contar con una herramienta que permita integrar y centralizar la información, para realizar análisis rápidos y eficientes

2.4.1. Definición de DataWarehouse

En toda empresa, día a día se genera gran cantidad de información, mucha de la cual se encuentra almacenada en las bases de datos que interactúan con los sistemas operacionales de las empresas. Un reto en la actualidad es lograr un correcto manejo de estos datos para generar informes útiles, es decir



claros concisos y en línea para que los altos directivos de la empresa puedan utilizarlos para su análisis y posterior toma de decisiones.

Según: Inmon (2002) “Un DataWarehouse es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones” p.54. Por otro lado, Kimball (1996) define al DataWarehouse como una copia de los datos transaccionales específicamente estructurada para la consulta y análisis.

La diferencia entre Inmon y Kimball, radica en la manera de construir el DataWarehouse, puesto que para Inmon la construcción inicia con el DataWarehouse para luego llegar hasta las áreas específicas o Datamarts (Top-down) y Kimball inicia con la construcción de los Datamarts para llegar al DataWarehouse (Bottom-up) (Rivadera, 2010).

Por lo antes expuesto, se puede concluir que un DataWarehouse es un repositorio de datos que se construye, nunca puede ser comprado, y es propio para cada empresa. Así también se puede concluir que un DataWarehouse no es un backup de todos los datos originados de los sistemas transaccionales, sino que se almacenan los datos necesarios después de pasar por un proceso de limpieza de manera que no existan datos erróneos en el almacén de datos.

2.4.2. Definición de Datamart

Un Datamart es un subconjunto de la información de un DataWarehouse enfocado en una sola área funcional o departamento de la empresa (Bernabeu, 2010). Al tratarse de los datos enfocados un área funcional específica de la empresa se puede restringir el acceso solo a un grupo de usuarios o a todo el personal del departamento.

2.4.3. Modelo Multidimensional

Según Kimball (1996), el modelo multidimensional es una técnica de diseño, cuyo objetivo es presentar los datos de forma simple y fácil de entender para el usuario. La creación de un modelo



multidimensional es un proceso iterativo y dinámico que consiste en: elegir el proceso de negocio para establecer el nivel de detalle con los que se manejará los datos en el DataWarehouse (nivel de granularidad), luego se eligen las dimensiones y se establecen los indicadores y las tablas de hechos (Rivadera, 2010).

Hechos

Los hechos son todos aquellos indicadores o medidas del negocio, que la empresa necesita obtener y analizar. Son todas las medidas numéricas que pueden ser obtenidas para futuros análisis en la empresa. Para establecer los hechos se debe conocer de manera exacta cómo calcularlos para evitar errores en los análisis.

Dimensiones

Las dimensiones almacenan información relativa a los factores que se pueden analizar en un hecho. Según: Chinchilla (2011), las dimensiones son las variables sobre las cuales se califican los datos. Las dimensiones tienen atributos que establecen la calidad del DataWarehouse. La dimensión tiempo es la dimensión más importante para el modelo dimensional, puesto que los hechos son almacenados según como se comporta a lo largo del tiempo.

Se concluye que, las tablas de hechos contienen información de medidas y las tablas de dimensiones contienen los metadatos sobre los hechos.

Tipos de modelos dimensional

Al realizar un modelo dimensional este puede variar, dependiendo de la forma de representar la jerarquía de las dimensiones y el nivel de granularidad que se va a aplicar al modelo.

Modelo estrella

El modelo estrella forma gráficamente un diagrama tipo estrella, de ahí deriva su nombre, está formado por una tabla central que es la tabla de hechos y de varias tablas de dimensiones que están relacionadas con la tabla central, a través de sus respectivas claves (Bernabeu 2010). El modelo estrella optimiza tiempos de respuesta de las consultas de los usuarios debido a que no se encuentra normalizado y por lo tanto se evitan las uniones entre tablas al momento de realizar las consultas. La Figura 3 es un ejemplo de la representación gráfica de un modelo de este tipo

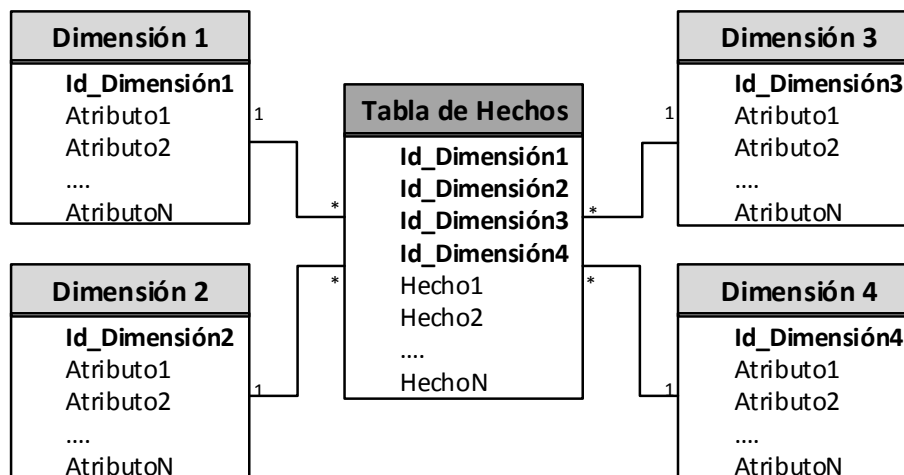


Figura 3: Modelo Estrella.

Fuente: Elaboración propia basada en: Bernabeu, D. (2010). *Esquema en Estrella*. [Figura]. En *Hefesto* (p. 37).

Modelo copo de nieve

Este tipo de modelo es una variación del modelo estrella, debido a que existe una tabla central que es la tabla de hechos que está relacionada con las tablas de dimensiones, pero que éstas a su vez pueden estar relacionadas con otras tablas de dimensiones (Bernabeu, 2010). De lo dicho anteriormente se puede ver que este modelo realiza una normalización de las tablas de dimensiones. La Figura 4 muestra gráficamente un ejemplo generalizado de un modelo de tipo copo de nieve.

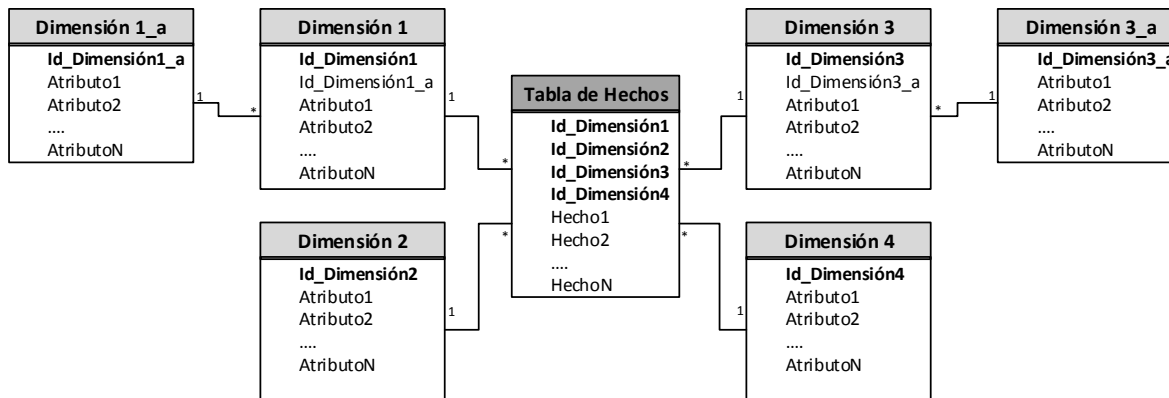


Figura 4: Modelo copo de nieve.

Fuente: Elaboración propia basada en: Bernabeu, D. (2010). *Esquema Copo de Nieve* [Figura]. En *Hefesto* (p. 39).

Modelo constelación

Este modelo es la integración de los modelos en estrella, es decir, cuando las tablas de hechos comparten dimensiones el modelo se convierte en un modelo constelación compuesto de varios modelos estrella (Bernabeu, 2010).

2.4.4. Metodología Hefesto

La metodología Hefesto explica los pasos que se deben seguir para la construcción de un DataWarehouse. Esta metodología fue creada por Darío Bernabeu, e inicia con el establecimiento de requerimientos y necesidades de información de los usuarios para confeccionar un esquema lógico realizando los respectivos procesos de extracción, transformación y carga de datos (ETLs) (Bernabeu, 2010). La Figura 5, indica gráficamente cada uno de los pasos de la metodología Hefesto. El primer paso de Hefesto genera 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 utilizará y se definen las tablas de hechos, 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. Es importante recalcar que la metodología Hefesto es independiente de las herramientas de software, por lo tanto, los 4 pasos definidos se pueden aplicar a cualquier problema sin importar si se cuenta con el software.

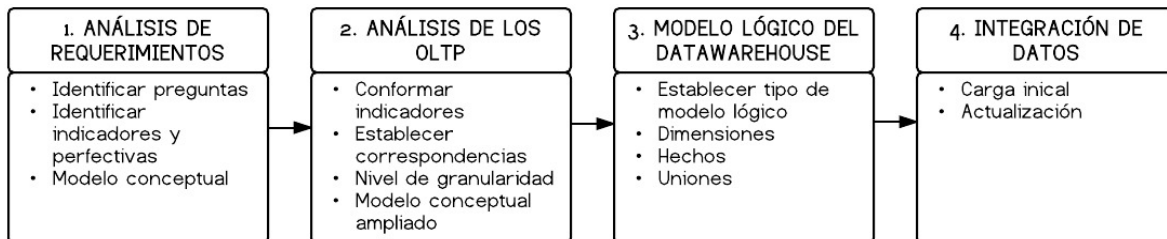


Figura 5: Pasos de la metodología Hefesto.

Fuente: Elaboración propia

A continuación, se explica cada uno de los pasos de la metodología:

Análisis de requerimientos

Esta es la fase inicial de la metodología, la cual consiste en identificar los requerimientos de los usuarios en base a diferentes técnicas, el resultado será un conjunto de preguntas que serán analizadas en el contexto de los objetivos de la empresa que generalmente están definidos en el Plan Estratégico.

Identificar preguntas

Según Bernabeu (2010) para identificar las preguntas se trabaja directamente con los usuarios, esta fase es muy importante puesto que esta será el soporte para los siguientes pasos de la metodología. Las preguntas identificadas incluirán variables de análisis con el fin de estudiar la información desde diferentes puntos de vista y perspectivas.

Identificar indicadores y perspectivas

Para la identificación de los indicadores y perspectivas se parte de las preguntas definidas anteriormente, puesto que estas deben ser descompuestas para obtener los indicadores y las

perspectivas que permitirán analizar dichos indicadores. Los indicadores serán valores numéricos y las perspectivas son entidades que examinarán los indicadores.

Modelo Conceptual

El siguiente paso en esta metodología es elaborar el modelo conceptual, el cual permitirá comprender de manera fácil y sencilla cuáles son los resultados deseados, con qué variables serán analizados y cuál es la relación entre estos. En la Figura 6 se puede visualizar gráficamente un ejemplo de un modelo conceptual basado en los indicadores y perspectivas.

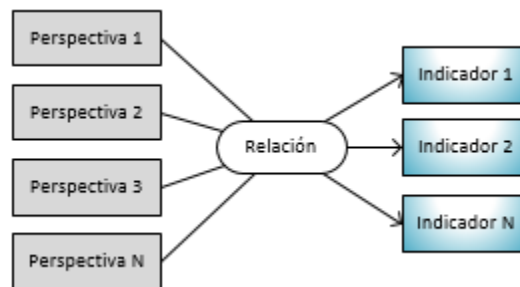


Figura 6: Ejemplo de modelo conceptual

Fuente: Elaboración propia basada en: Bernabeu, D. (2010). *Modelo Conceptual*. [Figura]. En *Hefesto* (p. 92).

Análisis de los OLTP

Esta fase consiste en analizar las fuentes OLTP para establecer la forma de cálculo de los indicadores, las correspondencias entre el modelo conceptual y sus respectivas fuentes de datos, además se definirá los campos necesarios para cada perspectiva, finalmente se realizará un modelo conceptual ampliado con toda la información resultante de esta fase.

Conformar indicadores

Para conformar cada indicador se debe establecer su fórmula de cálculo definiendo los hechos y las funciones de sumarización que se utilizarán para su cálculo.

Establecer correspondencias

En este paso se analizan las fuentes de datos existentes que se utilizarán para obtener la información necesaria para cada una de las perspectivas y los indicadores establecidos anteriormente en el modelo conceptual. Es importante que todos los elementos del modelo estén soportados en los OLTP. Por lo tanto, en este paso se decide si ciertas preguntas son parte del proceso o no, puesto que, si el usuario define preguntas que no pueden ser respondidas con los datos existentes en los sistemas informáticos de la empresa, estas son eliminadas de manera inmediata del proceso.

Nivel de granularidad

Luego de establecer las correspondencias un paso muy importante es establecer cada uno de los campos para cada una de las perspectivas; estos campos servirán para examinar y filtrar los indicadores del modelo planteado.

Modelo conceptual ampliado

Este paso consiste en ampliar el modelo conceptual establecido en pasos anteriores, y para ello se indicará en cada perspectiva los campos establecidos y en cada indicador su fórmula de cálculo. La Figura 7 indica un ejemplo de un modelo conceptual ampliado

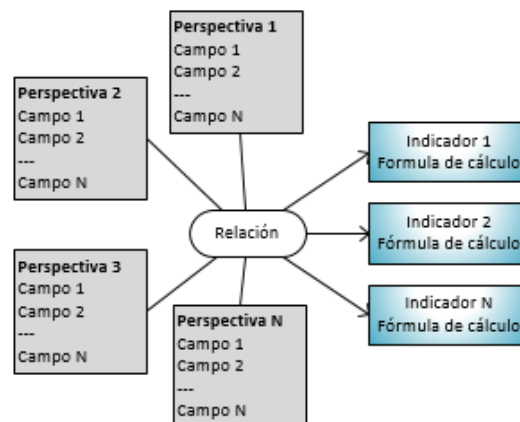


Figura 7: Ejemplo de modelo conceptual ampliado

Fuente: Elaboración propia basada en: Bernabeu, D. (2010). *Modelo Conceptual ampliado*. [Figura]. En *Hefesto* (p. 98).



Modelo lógico del DataWarehouse

Para construir el modelo lógico se parte del modelo conceptual ampliado, estableciendo el tipo de modelo a utilizar y luego diseñar las tablas de dimensiones y de hechos para finalizar realizando las uniones entre las tablas.

Tipo de modelo lógico del DataWarehouse

En este paso se selecciona el tipo de esquema que se utilizará para construir el DataWarehouse de la empresa, se escogerá el que más se adapte a las necesidades y requerimientos de los usuarios. En este punto se define entre los esquemas estrella, constelación o copo de nieve.

Tablas de dimensiones

En este paso se diseñan las tablas de dimensiones, esta actividad se realiza partiendo de las perspectivas definidas en el modelo conceptual, cada perspectiva corresponderá a una dimensión del DataWarehouse, para cada dimensión se definirá un nombre y una clave principal además de cada uno de los campos de la tabla.

Tablas de hechos

En este paso se definen las tablas de hechos que van a conformar los indicadores para el DataWarehouse. Una tabla de hechos debe tener un nombre, una clave principal que está compuesta de las claves principales de las dimensiones que están relacionadas, cada hecho definido será un campo de la tabla.

Uniones

Este paso consiste simplemente en establecer las uniones entre las tablas de hechos y las tablas de dimensiones.



Integración de datos

Luego de diseñar el modelo lógico el siguiente paso será poblarlo con los datos a través de procesos ETL y la carga de datos, se debe también definir políticas de actualización de los datos.

Carga inicial

La carga inicial de datos consiste en poblar el modelo de datos definido en el paso anterior, para lo cual se deberá diseñar los procesos ETL, esto se facilita a través del uso de software existente que facilita esta actividad y no se vuelva una tarea compleja. Los datos que se cargan en el DataWarehouse deben ser limpios, sin faltantes ni anómalos, primero se deben cargar los datos de las dimensiones para luego poblar las tablas de hechos.

Actualización

Luego de realizar la carga inicial de los datos en el DataWarehouse, es necesario establecer políticas para realizar la actualización de los datos, de tal manera que el almacén de datos siempre se mantenga actualizado con la información necesaria para la ayuda a la toma de decisiones.

2.4.5. Explotación de la información

Una vez que se ha construido un DataWarehouse, los usuarios podrán utilizar la información a través de herramientas de consulta y análisis, estas herramientas serán el vínculo entre el DataWarehouse y el usuario. Existen diferentes herramientas para que los usuarios puedan hacer uso de la información almacenada en el DataWarehouse. Según Bernabeu, 2010 las herramientas para consultas y análisis en general deben tener las siguientes características:

- Accesibilidad de la información de forma transparente para el usuario para que sólo se enfoquen en el análisis y no en el origen de los datos.



- Apoyo para la toma de decisiones
- Orientación a los usuarios finales, de tal manera que puedan analizar la información sin necesidad de tener conocimientos técnicos.

A continuación, se explica brevemente algunas de ellas:

Reportes y consultas

Son herramientas que permiten la elaboración de informes sobre los registros almacenados en el DataWarehouse. Estas herramientas permiten presentar los informes en diferentes formatos según las necesidades del usuario, por ejemplo, se puede incluir imágenes, enlaces a otros informes, etc.

OLAP

El procesamiento de análisis de datos es la herramienta más poderosa de un DataWarehouse, cuyo objetivo principal es brindar respuestas rápidas a preguntas complejas para que los usuarios puedan interpretar la información y tomar decisiones. Este tipo de herramientas permite el análisis de una gran cantidad de información, permitiendo que el usuario pueda encontrar mejores respuestas.

Dashboards

Este tipo de herramientas permite construir los informes de una forma visual muy llamativa, a través de reportes, consultas y gráficos de un tema en particular, pueden presentar la información en forma resumida de tal manera que el usuario pueda realizar un análisis de forma rápida.

Minería de datos (Data Mining)

Como se vio anteriormente la minería de datos es el componente de la Inteligencia de Negocios que permite analizar la información almacenada en el DataWarehouse. Estas herramientas permiten extraer el conocimiento a partir de la información histórica de una empresa, puesto que permite



inferir, predecir o estimar variables o comportamientos futuros. Esto permitirá la toma de decisiones basadas en el conocimiento de la información.

EIS (Executive Information System)

Este tipo de herramientas permiten la elaboración de informes resumen e indicadores clave para la gestión en una empresa, puesto que los usuarios podrán analizar la información en forma rápida y efectiva porque se trata de información simplificada y consolidada.

2.5. Metodologías de desarrollo de proyectos de software

El éxito de un proyecto mucha de las veces depende de una buena gestión y metodología de desarrollo del proyecto, es por esto que es muy importante saber gestionar y planificar de manera correcta los pasos o fases para el buen desarrollo del proyecto, de manera que estas puedan llegar a un feliz término cumpliendo con las necesidades identificadas a un inicio. Existen dos tipos de metodologías de desarrollo de proyectos informáticos: las tradicionales y las ágiles.

2.5.1. Metodologías tradicionales

En una metodología tradicional se planifica absolutamente todo el trabajo para poder empezar con el ciclo de desarrollo del producto. Una metodología tradicional sigue rigurosamente un ciclo de desarrollo y el producto es entregado al cliente únicamente cuando se ha terminado por completo en base a las necesidades identificadas durante la fase inicial del proyecto.

El ciclo de vida de una metodología tradicional organiza el proyecto en etapas o fases que se realizan una sola vez durante todo el proyecto y su ciclo de vida se trata de un modelo en cascada. Las fases definidas para los modelos tradicionales son: análisis, diseño, desarrollo, pruebas, implementación y mantenimiento. Una fase iniciará siempre y cuando se termine la anterior. El usuario no participa durante el desarrollo del proyecto únicamente recibe el producto final por tanto no existe validación por parte del usuario durante todo el ciclo de vida del proyecto.



2.5.2. Metodologías ágiles

La característica principal de este tipo de metodologías es que su ciclo de vida es iterativo e incremental con la participación activa del usuario a lo largo de todo el proyecto, quien puede validar en cada iteración, si el producto se está desarrollando de acuerdo a sus necesidades, evitando de esta manera los malos entendidos, ahorrando tiempo y dinero. Las metodologías ágiles permiten entregas parciales pero que son funcionales del producto permitiendo que en cada iteración mejorar la calidad del producto, asegurando que el producto final será lo que realmente el usuario necesita. Estas metodologías ponen de relevancia que la capacidad de respuesta a un cambio es más importante que el seguimiento estricto de un plan. Este tipo de metodologías se proponen porque para muchos clientes esta flexibilidad será una ventaja competitiva y porque estar preparados para el cambio significar reducir su coste (Figueroa, Solis, & Cabrera, 2008).

Metodología SCRUM

Entre las metodologías ágiles más conocidas se encuentra SCRUM, esta es una metodología ágil que permite establecer iteraciones cortas de desarrollo, denominadas sprints (Scrum.org, s.f.). Cada sprint es planificado 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 8, 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.

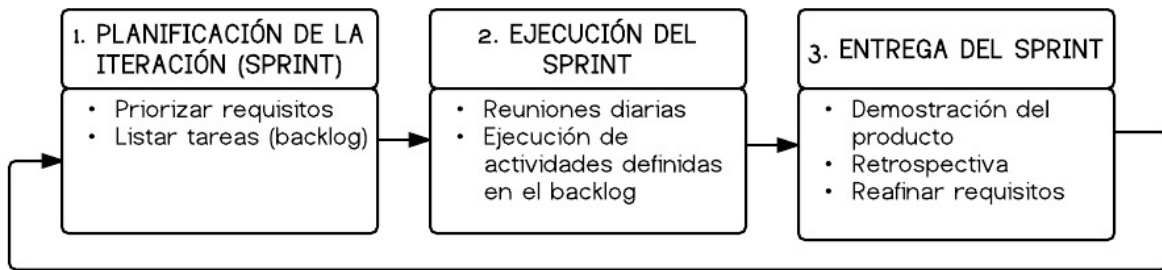


Figura 8: Metodología ágil de desarrollo de proyectos SCRUM

Fuente: Elaboración propia

2.6. Modelamiento de procesos de negocio utilizando notación BPMN

Un proceso del negocio no es más que un grupo de actividades que se realizan en una cierta secuencia sobre una o más entradas para producir una salida (White & Miers, 2010). El modelamiento de los procesos permite establecer el flujo de trabajo, de manera que cualquier usuario pueda comprender fácilmente las interrelaciones existentes entre las distintas actividades del negocio.

BPMN, es una herramienta gráfica desarrollada para modelar procesos del negocio, provee una notación estándar y comprensible para todo tipo de usuario, aun si no es un técnico. La notación BPMN utiliza un conjunto de elementos gráficos que están agrupados en categorías tales como: objetos de flujo, objetos de conexión, canales, artefactos y datos (Gómez, 2014). Para BPMN un proceso representa lo que una organización realiza (su trabajo) para lograr cumplir su propósito u objetivo. Un buen modelo debe ser: selectivo para representar los aspectos más relevantes de un proceso, debe ser exacto es decir se debe modelar la situación actual y no una suposición, además un buen modelo será comprensible y completo (White & Miers, 2010).

En la Figura 9, se pueden visualizar algunos ejemplos de los elementos utilizados para la notación BPMN.






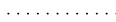






Objetos de flujo	 Tarea	 Subproceso	 Eventos	 Puerta de enlace
Objetos de conexión	 Flujo de secuencia	 Asociaciones	 Flujo de mensajes	
Objetos de datos	 Objeto de datos	 Almacén de datos	 Mensaje	
Artefactos	 Anotaciones	 Grupo		

Figura 9: Ejemplos de elementos de la Notación de BPMN

Fuente: Elaboración propia

2.7. 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 para examinar a sus deudores y cartera vencida (Lozada & Cruz, 2014). También se vio la utilidad de la integración de la información a través de un DataWarehouse para una institución sanitaria, en donde se considera fundamental dotar de un sistema para la toma de decisiones para los directivos (Castillo, 2014). En la mayoría de 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 el campo de 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 Luján, 2012), este proyecto señala que trabajar con un DataWarehouse simplifica procesos para tomar decisiones, puesto que el acceso es fácil y es beneficioso para los usuarios para que puedan hacer sus



propias consultas basados en los modelos multidimensionales. Se revisó también el análisis del rol del DataWarehouse en la educación superior (Guan, Nunez, & Welsh, 2002), 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 trabajo analizado de la Universidad de Phoenix: “*Driving decisions through academic analytics* (Pirani & Albrecht, 2005), indica que todas las Universidades buscan tener una herramienta que les permita realizar el análisis de información relevante para la toma de decisiones acertadas. Este trabajo concluye 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.

Referente a trabajos que utilizan SCRUM para desarrollar proyectos de construcción e implementación de un DataWarehouse , (Factos, 2014) indica que SCRUM se aplica para entregar vistas analíticas en corto tiempo para lo cual los Datamarts se desarrollan incrementalmente, realizando los ETLs sólo de un área, para la cual se crean los cubos OLAP, luego los informes, los cuadros de mando y se valida con los usuarios relacionados con el área involucrada, lo que se convierte en una ventaja porque se interacciona con el usuario final y se realiza rápidamente las correcciones necesarias para obtener el producto final, en este trabajo se concluye que SCRUM puede entregar resultados tangibles de manera progresiva tomando en cuenta que los usuarios puede cambiar constantemente los requerimientos (Reimundo, Nasimba, & Cueva, s.f.). Así también en un proyecto cuyo objetivo es construir un DataWarehouse para proporcionar información sobre el sistema financiero del Ecuador, se utilizó la metodología SCRUM para su desarrollo, realizando entregas funcionales de manera incremental en base a iteraciones y revisiones que permitieron la corrección de errores tempranamente y de esta manera entregar los resultados que cumplen los requerimientos de los usuarios (Factos, 2014). Otro caso analizado fue el trabajo “*An Action Research study on the use of Scrum to provide agility in Data Warehouse development*”, en éste evalúa la implementación de SCRUM en la construcción de un DataWarehouse para una empresa financiera, sugiere que para que la metodología sea exitosa debe haber apoyo total por parte de los directivos de la empresa, y debe darse la importancia necesaria al equipo de desarrollo organizado y dinámico con un líder que será el encargado de monitorear el equipo y establecer las reglas y métricas durante el desarrollo del proyecto para obtener el ansiado éxito (Mulder, 2010). Con respecto a la combinación de SCRUM



con BPMN, se encontró el caso de desarrollo de un ERP que aplica SCRUM para el desarrollo del proyecto y BPMN para la gestión de requerimientos. Este trabajo concluye indicando que el uso de SCRUM fue el adecuado debido a las características de la metodología ágil y que la notación BPMN permitió reducir el número de iteraciones mejorando la comunicación entre el equipo de desarrollo y los interesados del sistema (Alanocca & Ccahuana, 2015). Es importante anotar que no se encontró un trabajo que combine la metodología Hefesto, SCRUM y BPMN, por lo tanto, el reto de esta tesis fue realizar la combinación de los tres artefactos para la construcción e implementación del DataWarehouse y Datamarts para la Universidad de Cuenca.

Por todo lo antes expuesto, se deduce 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, es por eso que en base a la experiencia de los trabajos analizados y de acuerdo a lo explicado anteriormente: para crear el DataWarehouse o Datamart se utiliza la metodología Hefesto; para desarrollar proyectos informáticos se utiliza SCRUM; finalmente para definir procesos se utiliza la nomenclatura de BPMN. Este trabajo en los capítulos posteriores demostrará como se logró integrar Hefesto, SCRUM y BPMN dentro de un proceso de creación de un DataWarehouse que permita obtener productos de manera inmediata a través de la aplicación de SCRUM y gestionados por los procesos definidos en BPMN.



Capítulo 3.

CONSTRUCCIÓN E IMPLEMENTACIÓN DEL DATAWAREHOUSE

En este capítulo se describe la arquitectura que fue utilizada para la construcción del DataWarehouse o un Datamart de un área específica de la Universidad de Cuenca. Debido al alcance de este trabajo, la arquitectura se aplicó únicamente para la construcción del Datamart correspondiente a la Comisión de Evaluación de la Universidad.

3.1. Arquitectura del DataWarehouse

La arquitectura planteada para la construcción del DataWarehouse y para cualquier Datamart combina: la metodología Hefesto para la construcción de un DataWarehouse, la metodología SCRUM para desarrollo de proyectos y la notación BPMN para el modelado de los procesos de negocio (específicamente para modelar los ETLs necesarios para el DataWarehouse).

La combinación de metodologías y notaciones se realizó con el objetivo de definir un proceso único de creación de DataWarehouse para la Universidad de Cuenca a través de la reutilización de las diferentes actividades, lo que conlleva a la obtención de un producto que esté acorde a las necesidades del usuario, puesto que aplicando SCRUM con la participación activa de los usuarios y el equipo de desarrollo permitió avanzar rápidamente, ahorrando tiempo y esfuerzos, detectando errores o cambios en los requerimientos durante la aplicación de la metodología Hefesto para la construcción del DataWarehouse, obteniendo versiones del Datamart por cada *sprint* realizado. Por otro lado, la notación BPMN al ser estándar y conocida permite leer rápidamente como se modelan cada uno de los procesos ETL para la construcción del DataWarehouse.

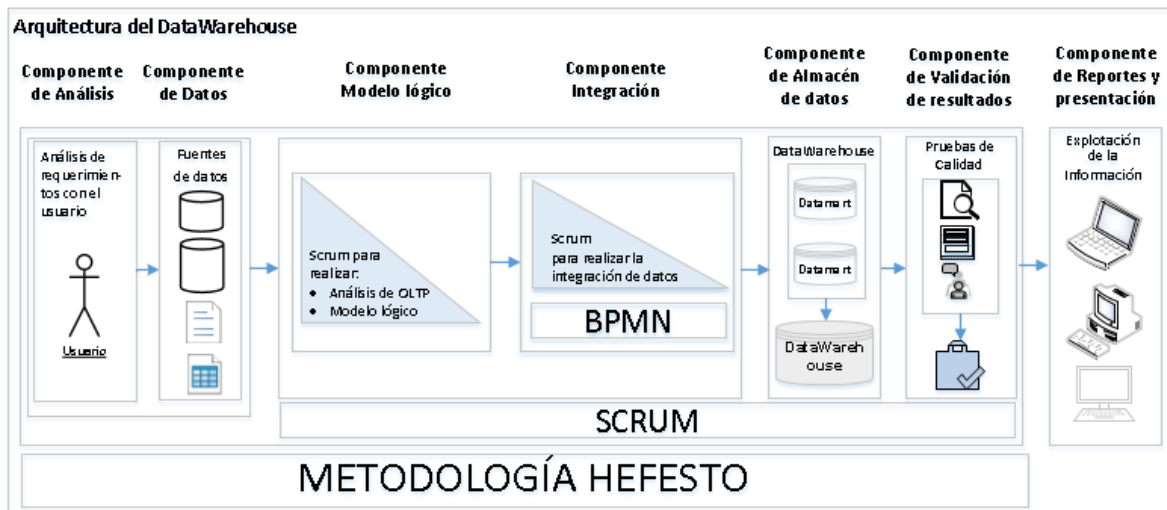


Figura 10: Arquitectura del Data Warehouse

Fuente: Elaboración propia

En la Figura 10 se visualiza la arquitectura propuesta en este trabajo con sus respectivos componentes, en donde se refleja el uso de Hefesto, SCRUM y BPMN. Como se puede observar en la figura, el primer paso de la aplicación de la metodología Hefesto corresponde al componente de análisis de requerimientos. El segundo y tercer paso de Hefesto abarca el componente de datos y el componente modelo lógico. La aplicación de SCRUM inicia en la segunda fase de la metodología Hefesto con una iteración para realizar el análisis de los OLTP y obtener el modelo lógico del Data Warehouse. El componente de integración corresponde al último paso de la metodología Hefesto, y también en este componente de la arquitectura de construcción del Data Warehouse a más de haber aplicado SCRUM con iteraciones para elaborar cada modelo multidimensional, se utilizó la notación BPMN para el modelado de los procesos ETL. El componente de almacén de datos y de validación también utiliza SCRUM para realizar todas las pruebas de validación del Data Warehouse. Finalmente está el componente de reportes y presentación, que es el que será utilizado por el usuario final para interactuar con el Data Warehouse.

Antes de iniciar con la aplicación de la metodología para la construcción del Data Warehouse, Hefesto recomienda que se realice un análisis de la empresa; por esta razón en este trabajo se debe conocer sobre la Universidad de Cuenca. En el Anexo 1. se puede revisar el análisis de la Universidad para comprender mejor cómo y porqué se realizó el diseño y la implementación del Data Warehouse.



A continuación, se detallan cada uno de los componentes que son parte de la arquitectura definida:

3.1.1. Componente de análisis

Este componente definido dentro de la arquitectura es una de las fases de la metodología Hefesto que permitió identificar los requerimientos de los usuarios, puesto que ellos son quienes dan la pauta del resultado que debe dar el DataWarehouse. Cada uno de los Datamarts fue definido luego de realizar entrevistas y encuestas a los usuarios de las áreas seleccionadas para la construcción del almacén de datos de la Universidad de Cuenca. Como resultado del análisis se obtuvo la Tabla 1, donde se listan las dependencias analizadas y los Datamarts con sus respectivos indicadores que en su mayoría pertenecen a la información solicitada por los criterios de evaluación, puesto que adicional a las entrevistas realizadas con los usuarios, se procedió también a revisar los criterios de evaluación establecidos por los organismos de control el Consejo de Educación Superior (CES, s.f.) y el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES, s.f.) para la evaluación y acreditación de las instituciones de nivel superior dando como resultado un listado de indicadores más representativo que servirá tanto para la toma de decisiones interna como para justificar los indicadores dentro de un proceso de evaluación de instituciones de educación superior. Los criterios establecidos por los organismos de control son: academia, eficiencia académica, investigación, organización e infraestructura.

Es importante anotar que en este componente se analizó que el DataWarehouse de la Universidad de Cuenca aportará significativamente para la consecución de algunos de los objetivos estratégicos que corresponden al eje “Gestión Institucional”, tales como: “Automatizar los procesos de la Universidad de Cuenca”, “Mejorar progresivamente la calidad de la prestación de los servicios informáticos” e “Implementar un sistema de planificación institucional”. Este último se puede especificar más, puesto que se lo puede relacionar con la acción estratégica: “Generar un sistema estadístico institucional”. Por último, también vale la pena recalcar que debido al alcance de este trabajo se puede obtener la información de uno de los indicadores definidos en el plan estratégico que es el “Porcentaje de docentes que superan el 80% de la evaluación, el cual corresponde al objetivo estratégico: “Mejorar la calidad de la docencia universitaria”, correspondiente al eje Docencia.



Tabla 1: Datamarts e indicadores establecidos para el DataWarehouse

Dependencias	Datamart	Indicadores
<ul style="list-style-type: none"> - Facultades - Consejo Académico - Direcciones de carrera 	Datamart 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 x Carrera en un periodo lectivo
		Estudiantes nuevos matriculados por carrera en un periodo lectivo
<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
	Datamart de Posgrados	Promedios de estudiantes por periodo lectivo y por programa de posgrado



- Dirección de Posgrados		Número de matriculados por programa de posgrado y por año
		Número de posgrados aprobados y en ejecución por año
- Dirección de Evaluación	Datamart de Proyectos	Número de participantes por proyecto por dependencia y por año
- Dirección de Planificación		Total de presupuesto por proyecto por dependencia y por año
- 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

Fuente: Elaboración propia

Como se estableció en el alcance de este trabajo, a continuación, se describen las preguntas identificadas para el Datamart de evaluación:

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 el número de participantes por facultad y género por cada proceso de evaluación. En otras palabras “Número de participantes en el proceso de evaluación, por facultad y género en cada proceso de evaluación”.

Luego de identificar las preguntas el siguiente paso fue identificar los indicadores que serán realmente útiles y las perspectivas que son los objetos mediante los cuales se evaluarán los indicadores. Continuando con el paso práctico de este trabajo se analizaron las preguntas identificadas para descomponer y visualizar los indicadores y las perspectivas.

que la relación que permite la unión entre las diferentes perspectivas y los indicadores requeridos se llama “Evaluación”

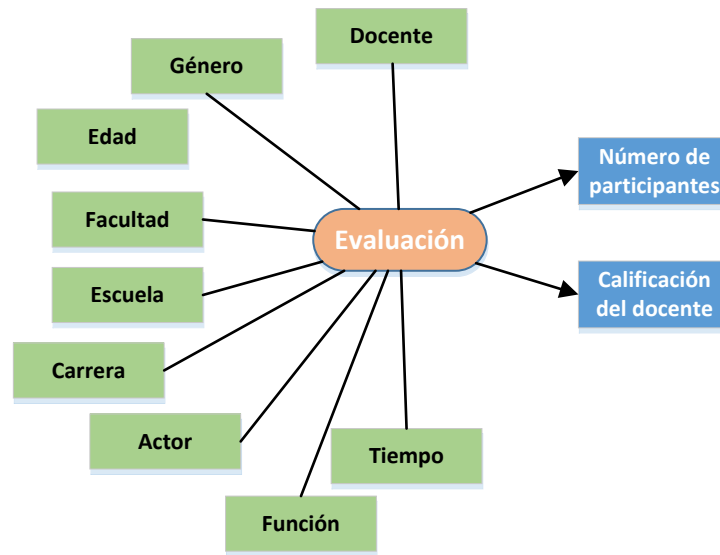


Figura 12: Modelo conceptual del Datamart de evaluación

Fuente: Elaboración propia

3.1.2. Componente de datos

En este componente se realizó un análisis de todas las fuentes de datos disponibles para la obtención de los datos necesarios para poblar cada uno de los Datamarts que serán parte del DataWarehouse. La Universidad de Cuenca almacena los datos en diferentes tipos de bases de datos, entre las cuales están Oracle 11g, DB2 sobre un AS400, Mysql y Postgresql, además se pudo constatar que en algunas dependencias y facultades todavía existe información en archivos de Excel. Para el caso de estudio de este trabajo (evaluación del docente), se detectó que los datos se almacenan en dos bases de datos distintas, la información correspondiente a los datos académicos de los docentes, estudiantes y sus respectivas asignaturas se almacenan en la base de datos Oracle 11g y lo referente a los datos personales de los docentes se almacena en una base de datos DB2 sobre un servidor AS400. En el Anexo 2, se puede observar el diseño entidad-relación de la base de datos que se utiliza en el sistema de gestión de evaluación del desempeño del docente, conocido en la Universidad como SGE. En la

Figura 13 se puede observar un resumen de cómo están organizadas las fuentes de datos para el Datamart de evaluación

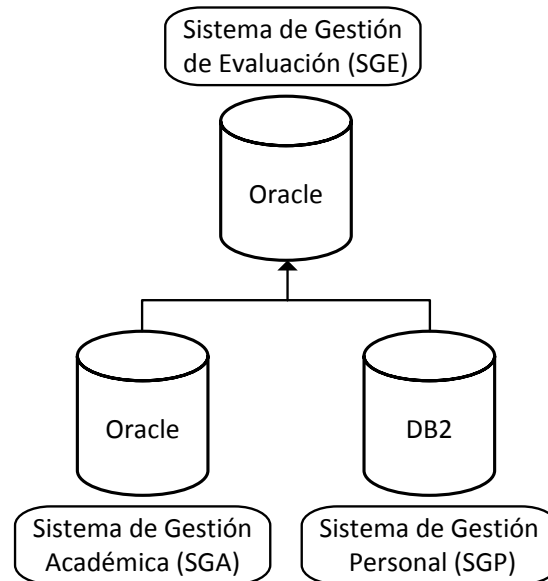


Figura 13: Fuentes de datos para el Datamart de evaluación

Fuente: Elaboración propia

Basado en los indicadores y perspectivas definidos en el componente de análisis se realizó una revisión exhaustiva de las bases de datos para identificar si es posible responder a las preguntas planteadas por los usuarios, concluyendo que efectivamente las fuentes contienen los datos suficientes para construir el Datamart de evaluación docente.

3.1.3. Componente Modelo Lógico

En este componente se utilizó la metodología ágil 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, las actividades realizadas en el sprint se resumen en la Tabla 3.



Tabla 3: Actividades planificadas para el sprint del componente modelo lógico

Actividades planificadas
1. Conformar indicadores y establecer forma de cálculo
2. Analizar diagrama entidad-relación y diccionario de datos
3. Establecer las correspondencias entre las fuentes de datos y el modelo conceptual
4. Definir nivel de granularidad
5. Definir el ámbito de la perspectiva tiempo
6. Elaborar modelo conceptual ampliado
7. Confeccionar el modelo lógico de la estructura del DataWarehouse
8. Realizar las uniones entre las dimensiones y los hechos

Fuente: Elaboración propia

Como se puede ver en la Tabla 3, la primera actividad del sprint fue conformar los indicadores, estableciendo la manera de cálculo de cada uno de ellos. A continuación, se explica cómo se obtuvieron y cuál fue la función de cálculo para obtener cada uno de los indicadores definidos en el modelo conceptual (Figura 12).

- Indicador: “Número de participantes”

Este indicador se obtiene al contar cuantos actores entre evaluadores y evaluados participaron en el proceso de evaluación.

- Hechos: Número de participantes
- Función de cálculo: COUNT

- Indicador: “Calificación del docente”

La calificación de los docentes (hecho) se calcula de manera diferente, dependiendo del proceso de evaluación:

- Hechos: Calificación del docente
- Función de cálculo: La función de cálculo depende del año del proceso, debido a que hasta antes del año 2014, la calificación de la evaluación del docente se calculaba de una forma diferente a la forma de cálculo de los procesos de evaluación a partir del año 2014, como se puede ver en la Tabla 4.

Tabla 4: Conformación de indicadores para el Datamart de evaluación

Procesos anteriores al 2014	Procesos a partir del 2014
$\sum \text{Calificaciones del docente según tipo de actor}$	$\frac{(D \times H.D) + (I \times H.I) + (D.G \times H.D.G)}{\text{Dedicación docente}}$
<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>	

Fuente: Elaboración propia

La siguiente actividad del sprint fue analizar el modelo entidad-relación y establecer las correspondencias entre las fuentes de datos y el modelo conceptual definido en el componente de análisis. El resultado de esta actividad se lo puede observar en la Figura 14, en donde se refleja la correspondencia entre el modelo entidad-relación visualizado en el Anexo 2 y el modelo conceptual definido en el componente de análisis.

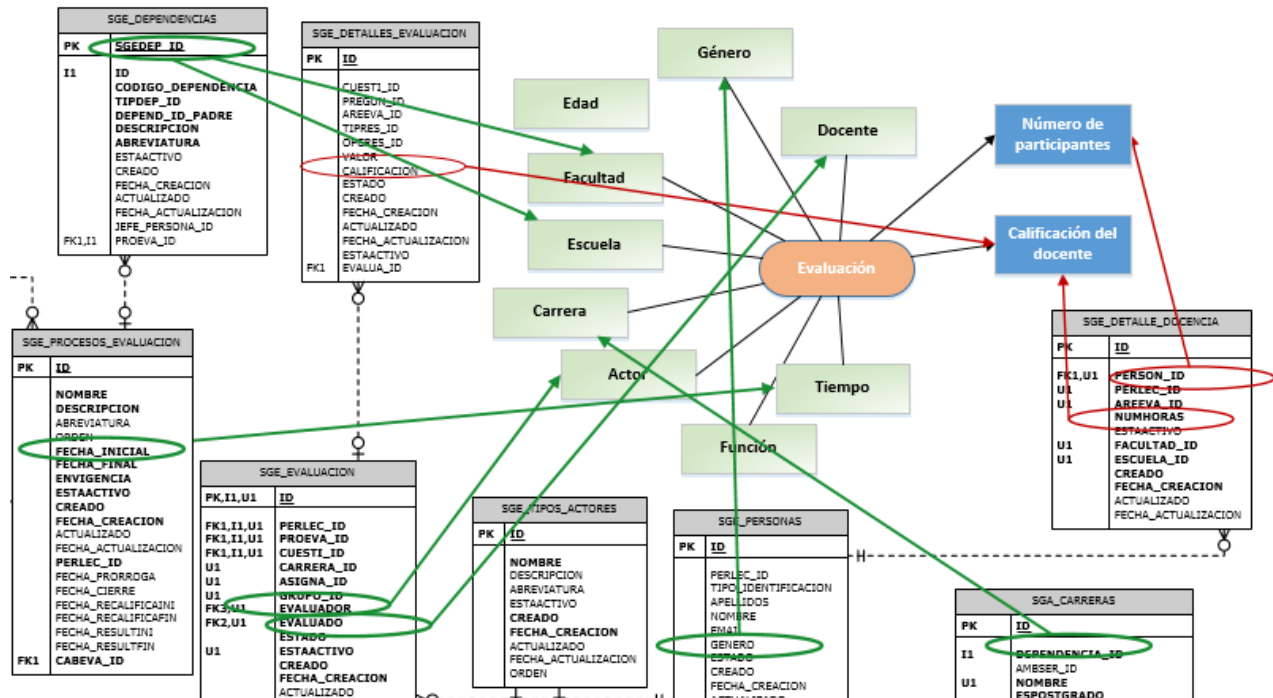


Figura 14: Correspondencias para el Datamart de evaluación

Fuente: Elaboración propia



Luego de establecer las correspondencias, se estableció el nivel de granularidad, definiendo los campos más relevantes en cada una de las perspectivas que fueron identificadas para filtrar los indicadores. En el caso específico de la evaluación del docente, para definir los campos de cada perspectiva, primero se analizó el diccionario entidad-relación y el diccionario de datos, que se puede observar en el Anexo 3. En la Tabla 5, se puede visualizar como se concibieron los campos para cada una de las perspectivas. 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 5: Nivel de granularidad para las perspectivas del Datamart de evaluación

Perspectiva	Campos	Diagrama entidad-relación
Docente	person_id apellidos nombres edad	adminuc.personas_naturales
Género	descripción	adminuc.personas_naturales
Edad	Edad	adminuc.personas_naturales
Facultad	descripción	adminuc.dependencias
Escuela	descripción	adminuc.dependencias
Carrera	Nombre	academico.sga_carreras
Actor	Nombre	evaluacion.sge_tipos_actores
Función	descripcion	evaluacion.sge_areas_evaluacion
Tiempo	Anio	evaluacion.sge_procesos_evaluacion

Fuente: Elaboración propia

Después de establecer la granularidad para las perspectivas se elaboró el modelo conceptual ampliado, indicando gráficamente los campos de cada perspectiva y la fórmula de cálculo para cada indicador. La Figura 15, muestra cómo se estableció este modelo conceptual ampliado para el caso de estudio de este trabajo.

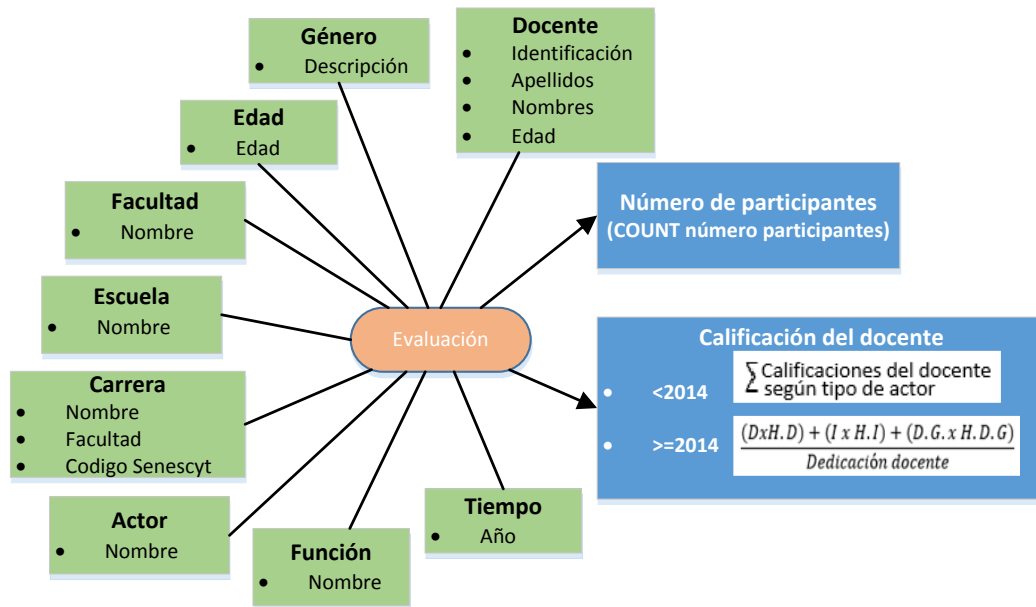


Figura 15: Modelo conceptual ampliado para el Datamart de evaluación

Fuente: Elaboración propia

Modelo Lógico

Como parte de este componente, se procedió a confeccionar el modelo lógico de la estructura del DataWarehouse en base al modelo conceptual ampliado levantado en el paso anterior. Primero se definió el tipo de esquema a utilizar, el mismo que para el caso de del Datamart de evaluación, al igual que para todos los demás se escogió el esquema estrella, elegido por sus características y ventajas (Bernabeu, 2010). A continuación, se procedió a diseñar las tablas de dimensiones y las tablas de hechos que conforman el Datamart de evaluación, tal y como se puede visualizar en la Figura 16, en la que se indica los campos que se incluyeron en cada tabla de dimensiones. Por otro lado, la Figura 17 representa la tabla de hechos que se definieron para este Datamart.

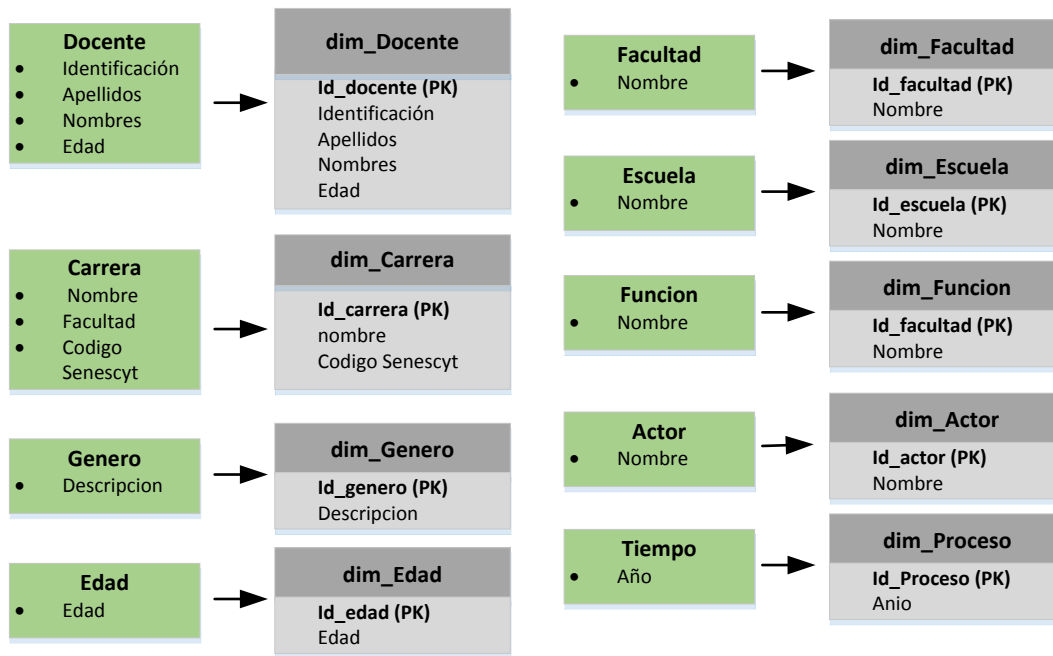


Figura 16: Tablas de dimensiones para el Datamart de evaluación

Fuente: Elaboración propia

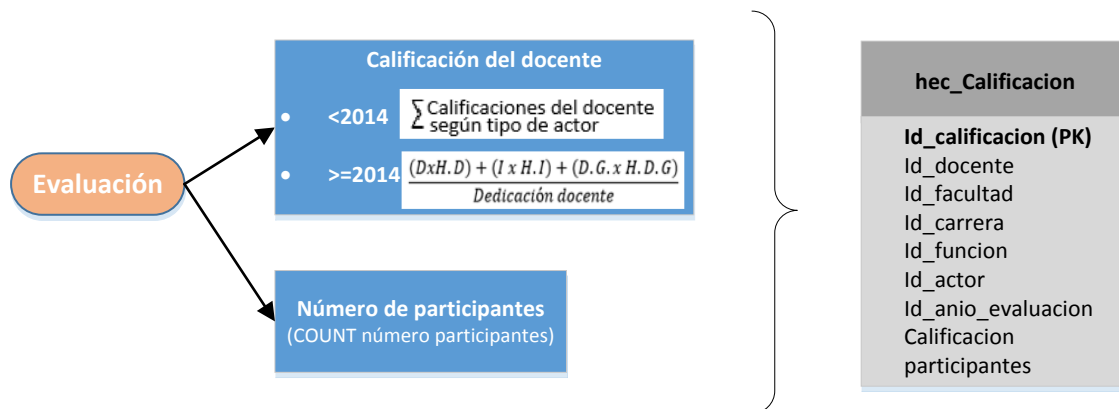


Figura 17: Tablas de hechos para el Datamart de evaluación

Fuente: Elaboración propia

Para finalizar las actividades del componente modelo lógico, se procedió a realizar las uniones necesarias entre las dimensiones y la tabla de hechos, tal y como se puede ver en la figura 18. Como se puede observar el modelo multidimensional está definido sobre un modelo entidad-relación.

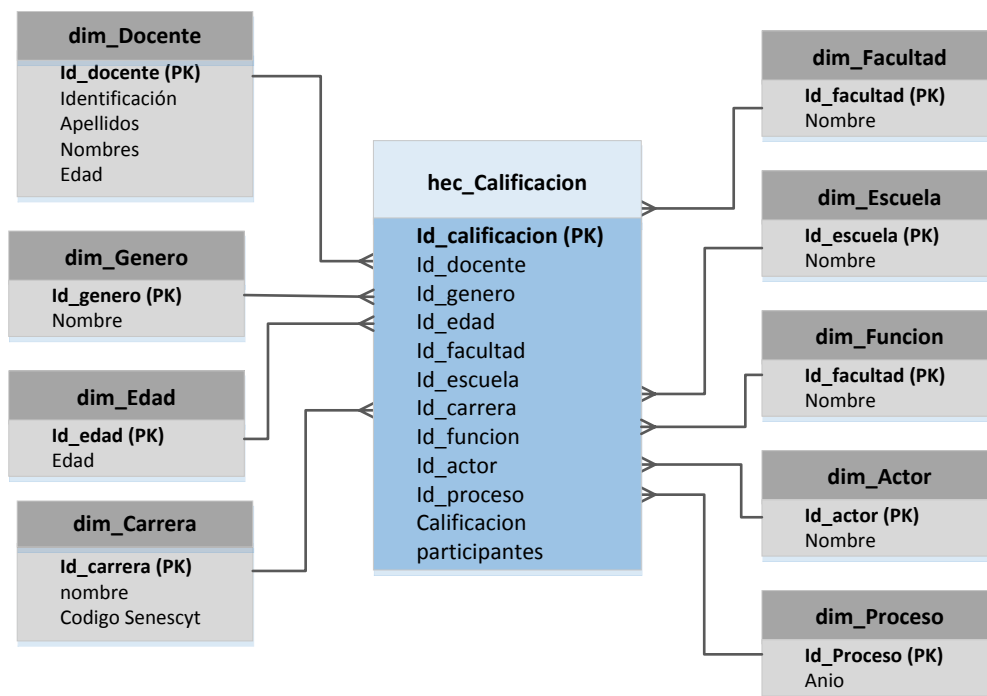


Figura 18: Uniones del Datamart de evaluación

Fuente: Elaboración propia

3.1.4. Componente Integración de datos

Una vez construido el modelo lógico, el objetivo principal del componente de integración es poblar el Datamart con los datos obtenidos de los sistemas operacionales, sin embargo, sobre estos datos se aplicaron técnicas de limpieza para evitar datos faltantes o anómalos en el DataWarehouse. En este componente también se aplica la metodología Hefesto combinado con SCRUM, sin embargo, se planificó un sprint para la construcción de cada proceso ETL, que permitió la población del almacén de datos. Por otro lado, también se debe notar que para modelar cada proceso ETL para la limpieza de información se utilizó la notación BPMN, con el objetivo de que los modelos resultantes sean entendidos por todos los usuarios. Las actividades realizadas en el sprint planificado para el componente de integración se resumen en la Tabla 6.



Tabla 6: Actividades planificadas para el sprint del componente de integración de datos

Actividades planificadas
1. Analizar la forma de realizar la limpieza y carga de los datos
2. Modelar los procesos ETL con BPMN
3. Implementar los procesos ETL con la herramienta Kettle
4. Crear los cubos multidimensionales
5. Pruebas unitarias y de integración
6. Pruebas de validación de datos
7. Pruebas de rendimiento
8. Pruebas de regresión
9. Elaborar e implementar los dashboards y reportes para el usuario

Fuente: Elaboración propia

A continuación, se detalla cómo se realizó la carga de datos para el Datamart de evaluación:

El sistema de gestión de evaluación inició su operación en el periodo lectivo marzo – agosto 2010, los procesos que fueron realizados a partir de esta fecha hasta el año 2013 aplicaron una forma de evaluación diferente a los procesos ejecutados desde el año 2014 en adelante, razón por la cual para obtener las calificaciones de los docentes se realizaron procesos de limpieza diferentes para cada caso. Otro asunto a tomar en cuenta durante el proceso de limpieza de datos fue que al inicio de las evaluaciones hubo carreras que se encontraban en otras dependencias que eran consideradas únicamente como programas y no como facultades, por lo cual también existió la necesidad de realizar un mapeo de estos programas con las facultades a las que ahora pertenecen las carreras. Entre las carreras que se encuentran en esta situación están: Cine y Audiovisuales y la carrera de Genero y Desarrollo, que ahora pertenecen a la Facultad de Filosofía y Jurisprudencia respectivamente. La actualización de los datos de evaluaciones es realizada después del cierre de cada proceso de evaluación, esto debido a que las calificaciones finales se obtienen únicamente luego de que se haya superado las recalificaciones y rectificaciones solicitadas por los docentes a la Comisión Interna de Evaluación.

Modelado de procesos ETL

El objetivo de un proceso ETL es seleccionar los datos para limpiarlos y transformarlos de manera que se pueda realizar la carga de los datos con el objetivo de que puedan ser analizados por los

usuarios. Todo ETL es un proceso que tiene entradas, operaciones y salidas, por esta razón se considera conveniente que sean modelados. En este trabajo se aplicó la notación BPMN para su modelamiento.

Cada sprint que se ejecute en este componente, dará como resultado un producto, estos son los cubos multidimensionales, de manera que, si el momento de realizar las revisiones con los usuarios se detectan errores o inconsistencias o incluso inconformidades, en un nuevo sprint se consideran estos desafíos para obtener mejores resultados.

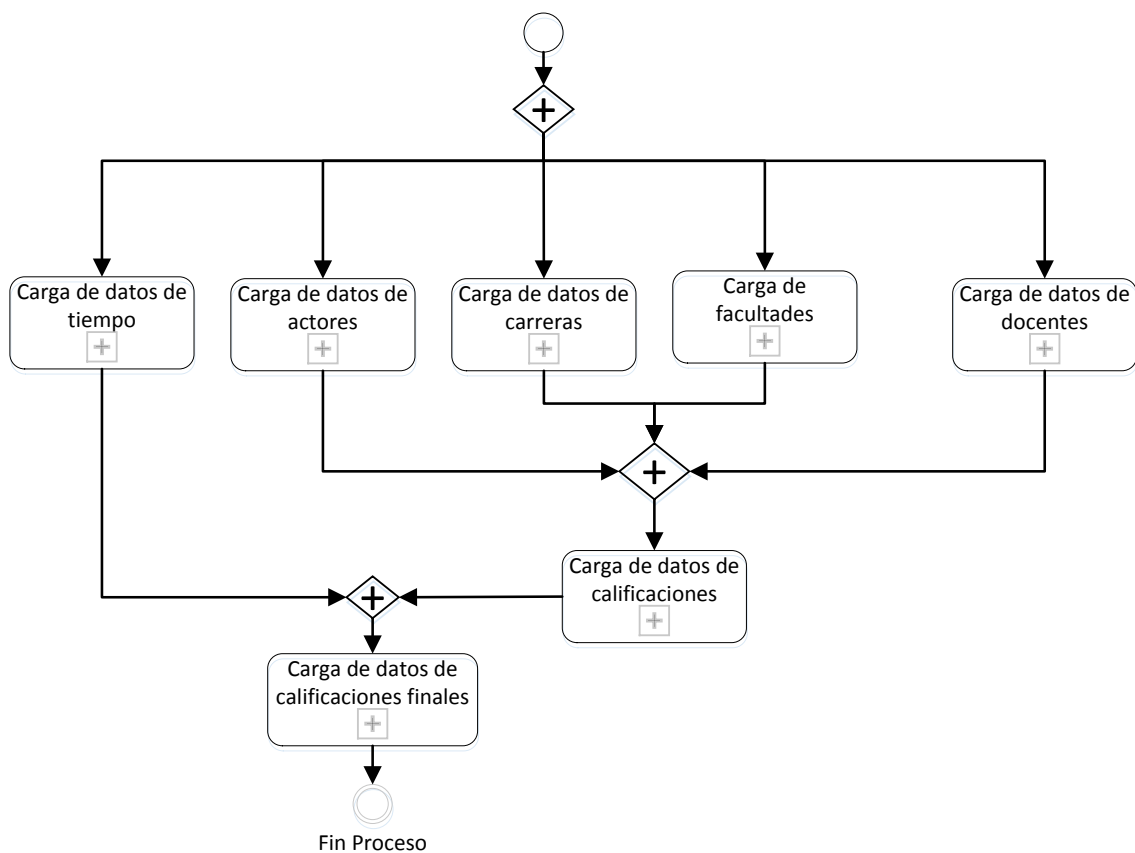


Figura 19: Vista general del proceso de carga de datos de evaluación

Fuente: Elaboración propia

Como se explicó anteriormente, para realizar el modelamiento de los procesos ETL, en este componente se hace uso de la notación BPMN. En la Figura 19 se puede observar una vista general sobre el modelado de la carga inicial de datos del Datamart de evaluación y en el Anexo 2, se puede

encontrar el modelado de todos los demás procesos ETL utilizados para la carga de datos del Datamart de Evaluación.

Una vez definidos los procesos ETL para la población del Datamart, se utilizó la herramienta Kettle, la cual es una herramienta de Pentaho Data Integration, que permite implementar los procesos de extracción, transformación y carga de datos (Naranjo & Sáenz, 2012). Para poblar el Datamart de evaluación se diferenciaron los procesos de evaluación realizados desde el 2010 hasta el año 2013 de los realizados desde el año 2014 en adelante, esto se realizó debido a que el cálculo de la calificación en los primeros procesos es diferente a los realizados desde el 2014 en adelante, además de la forma de cálculo, a partir del 2014 se toman en cuenta las carreras en las cuales dictan clases los docentes, puesto que anteriormente se mantenía la figura de escuela y el docente era calificado por el director de escuela a diferencia de los nuevos procesos en los cuales es el director de carrera quien realiza la calificación, sin embargo también se mantiene el concepto de escuela. Otro factor que interviene para realizar una diferencia es que en los nuevos procesos a más del decano y el director de carrera que son quienes califican al docente, también aparece un nuevo actor y se trata de los pares académicos nombrados por cada facultad. En la Figura 20 se puede visualizar los procesos de limpieza de los datos según los procesos antes mencionados, los demás procesos ETL ejecutados para el Datamart de evaluación se pueden en el Anexo 2.

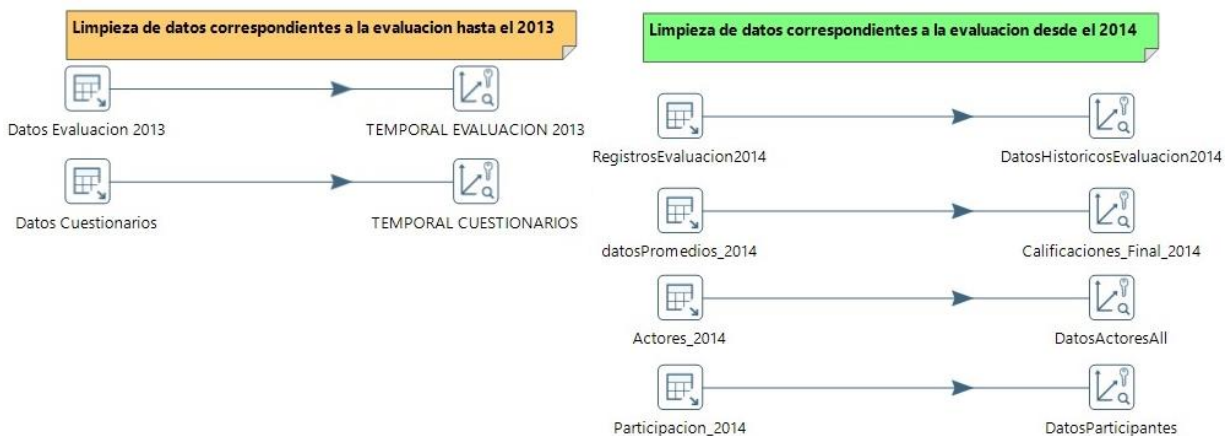


Figura 20: Procesos ETL para la limpieza de datos

Fuente: Elaboración propia



En la Figura 20 los datos iniciales para realizar la limpieza y transformación son obtenidos a través de consultas realizadas a las bases de datos que intervienen en el proceso y que fueron analizadas anteriormente.

Creación de cubos multidimensionales

A partir del modelo que se estableció en el componente modelo lógico, se elaboraron seis cubos multidimensionales, tres de ellos se crearon con la finalidad de obtener información sobre los procesos de evaluación realizados desde el año 2010 hasta el 2013, dos para los procesos realizados desde el año 2014 en adelante y un último cubo para obtener información sobre las calificaciones finales de los docentes por facultad, sin importar la carrera o escuela en la que dictaron clases. A continuación, se explican en forma más detallada cada uno de los cubos. Cabe recalcar que los cubos correspondientes a los procesos anteriores al año 2014 permiten filtrar la información por: año de evaluación, docente, genero, edad, facultad, escuela y actor; por otro lado, los procesos realizados a partir de año 2014 además de filtrar según los criterios nombrados anteriormente, también permite filtrar por carrera.

A continuación, se enumeran los cubos multidimensionales que se crearon en base al modelo lógico:

1. Cubo para obtener los datos de calificaciones obtenidas de procesos realizados desde el 2010 al 2013, en cada uno de los cuestionarios de la evaluación, en este cubo intervienen las dimensiones Tiempo, Docente, Genero, Edad, Facultad, Escuela, Actor y Cuestionario. En la aplicación se lo nombró como: 2013_evaluacion_cuestionarios.
2. Cubo para consultar las calificaciones de los docentes en los procesos realizados desde el año 2010 hasta el 2013. Intervienen las dimensiones Tiempo, Docente, Genero, Edad, Facultad y Escuela. Se lo denomina: Cubo 2013_evaluacion_docente.
3. El cubo que permite contestar la segunda pregunta planteada en este trabajo para los procesos de evaluación realizados hasta el año 2013 se lo llamó: Cubo 2013_evaluacion_funcion. Intervienen las dimensiones tiempo, docente, actor, función, género y edad.
4. El cubo para consultar información sobre la información de las evaluaciones que obtuvieron los docentes en los procesos de evaluación realizados a partir del año 2014 en adelante se denominó Cubo 2014_evaluacion_docente. Para este cubo se ha agregado la dimensión Carreras para

obtener la calificación del docente incluso por carrera; por lo tanto, este cubo sirve para contestar la primera pregunta planteada en este trabajo. Un cubo más creado para los procesos posteriores al 2013, se elaboró para obtener la calificación de los docentes en estos procesos también hasta el nivel de función y por actor. El cubo se llama: Cubo 2014_evaluacion_funcion.

5. El último cubo a más de contestar la tercera pregunta que consiste en obtener el número de participantes en las evaluaciones por cada proceso, también se puede obtener información e calificaciones finales que han obtenido los docentes en cualquiera de los procesos realizados desde que inicio el sistema de evaluación, esto debido a que para este cubo no se considera ni la escuela ni la carrera, únicamente la facultad.

En la Figura 21, se puede ver un ejemplo de la implementación de los cubos

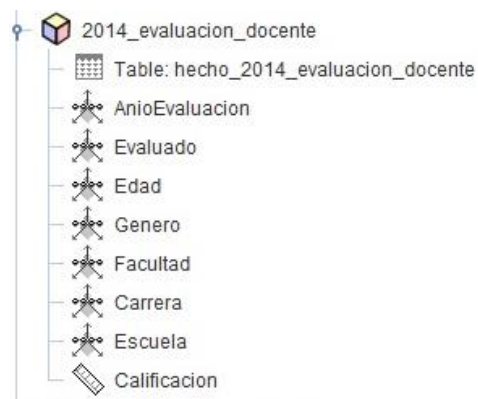


Figura 21: Ejemplo de un cubo multidimensional

Fuente: Elaboración propia

Es importante observar que el Datamart de evaluación, además de responder las preguntas planteadas en este trabajo, aporta para la consecución de los objetivos estratégicos analizados en el componente de análisis.

3.1.5. Componente almacén de datos

Este componente, se refiere al repositorio de datos, que como se vio anteriormente provienen de diferentes fuentes de datos de los sistemas operacionales de la Universidad. Este componente se logra luego de cumplir todas las actividades planificadas de los sprints planificados en el componente de

integración. El componente almacén de datos, está formado por el DataWarehouse empresarial el cual se trata de una vista general de la empresa, y por todos los datamart, los mismos que responden a datos de interés para áreas específicas como, por ejemplo, talento humano, dirección financiera, estudiantes, evaluación del docente, etc. Tal y como se estableció en el componente Modelo lógico, el proyecto utilizó el almacenamiento centralizado de datos 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.

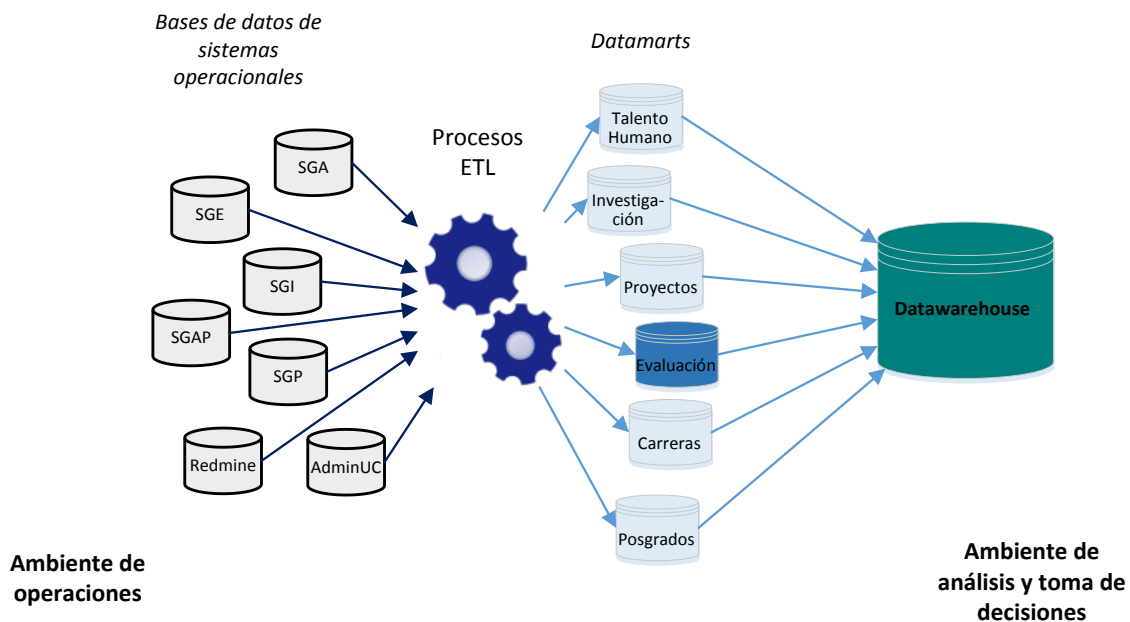


Figura 22: DataWarehouse de la Universidad de Cuenca

Fuente: Elaboración propia

El DataWarehouse es la herramienta que la Universidad utilizará para planificar, controlar y para la toma de decisiones relacionadas a cualquiera de las áreas que fueron incluidas para la construcción del almacén de datos, esto será posible puesto que la información es integrada y global. En la Figura 22 se puede observar gráficamente una visión general de la manera de cómo está organizado el DataWarehouse de la Universidad, incluyendo el Datamart de Evaluación.



3.1.6. Componente de validación de resultados

Este componente consiste en validar que el Datamart conteste las preguntas planteadas en el componente de análisis, que todos los resultados son útiles para los usuarios y que la solución implementada sea de calidad y aceptada por el usuario final. La explicación de este componente se hará más detalladamente en el capítulo cuatro, sin embargo, es importante mencionar que este componente también aplica SCRUM para realizar cada una de las pruebas planteadas.

3.1.7. Componente de reportes y presentación

El componente de reportes y presentación, es considerado como el más importante desde el punto de vista de los usuarios, puesto que, a través de éste el usuario interactúa con el DataWarehouse para obtener los reportes y datos necesarios para la toma de decisiones. Luego de revisar las características y bondades de algunas herramientas de inteligencia de negocios, tales como Bussines Object, Cognos, Microstrategy, Oracle BI, entre otros; se optó por Pentaho (Naranjo & Sáenz, 2012), teniendo en cuenta el costo de la herramienta, puesto que la versión Community que fue utilizada es gratis. Para que los usuarios puedan acceder a la herramienta se definieron distintos niveles de acceso, a través de la creación de roles que fueron asignados a los usuarios para revisar reportes o para realizar análisis de la información. Además, también se definió el acceso por 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.

En la Figura 23 se pueden ver algunos ejemplos de los reportes elaborados por la herramienta y que fueron desarrollados luego de que el usuario validara los resultados y se comprobó que el Datamart puede responder a las preguntas planteadas. En la Figura 23 se puede ver en forma gráfica la respuesta a la tercera pregunta sobre el número de participantes, así como también la respuesta a la primera pregunta que corresponde a las calificaciones de los docentes por facultad.

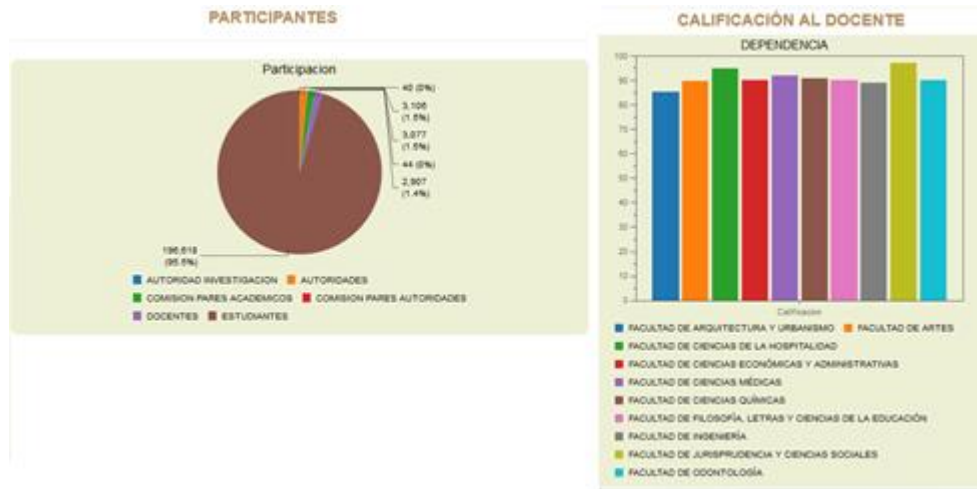


Figura 23: Ejemplos de reportes obtenidos a partir del Datamart de evaluación

Fuente: Elaboración propia

3.2. Aplicación de la metodología SCRUM

Como se mencionó anteriormente, para el desarrollo de este trabajo se aplicó la metodología SCRUM. En primer lugar, se definió el equipo de trabajo con tres integrantes, siendo uno de ellos el Scrum master, quien fue la persona que lideró al equipo y trabajó directamente con el dueño del producto. En vista de que el trabajo se centró en el Datamart de Evaluación, se estableció como Product Owner o dueño del producto al Director de la Comisión de Evaluación de la Universidad, quien participó muy de cerca en el proyecto y realizó las pruebas y validaciones de los resultados de cada sprint ejecutado.

Antes de iniciar la ejecución de los sprints, según la arquitectura definida en la sección 3.1, lo primero que se hizo fue definir los requerimientos que fueron plasmados a través de las tres preguntas definidas en la sección 3.1.1 (componente de análisis). Luego de realizar esta actividad se pudieron establecer el modelo conceptual y las fuentes de datos.

En la Tabla 7, se puede visualizar un resumen de cómo se ejecutaron las actividades para el primer sprint definidas en la sección 3.1.3



Tabla 7: Actividades ejecutadas en el primer sprint

Actividad planificada	Producto de la Actividad
1. Conformar indicadores y establecer forma de cálculo	El resultado de esta actividad se puede visualizar en la sección 3.1.3 (Tabla 4)
2. Analizar diagrama entidad-relación y diccionario de datos	Revisar sección 3.1.3. (Figura 14)
3. Establecer las correspondencias entre las fuentes de datos y el modelo conceptual	
4. Definir nivel de granularidad	Resultado en la sección 3.1.3. (Tabla 5)
5. Definir el ámbito de la perspectiva tiempo	
6. Elaborar modelo conceptual ampliado	El resultado se lo puede ver en la Figura 15 de la sección 3.1.3.
7. Confeccionar el modelo lógico de la estructura del DataWarehouse	Figuras 16 y 17 de la sección 3.1.3.
8. Realizar las uniones entre las dimensiones y los hechos	El resultado se lo puede ver en la Figura 18 de la sección 3.1.3.

Fuente: Elaboración propia

Luego de establecer el modelo lógico del DataWarehouse, se procedió a planificar y modelar los procesos ETL correspondientes al componente de Integración de datos. Una vez planificados los ETLs, se ejecutaron 5 sprints, uno por cada uno de los cubos definidos en la sección 3.1.4. En la Tabla 8 se puede ver un resumen de las actividades que fueron ejecutadas en cada sprint planificado.

Tabla 8: Actividades ejecutadas en cada sprint del componente de integración

Actividades planificadas	Producto de la Actividad
1. Analizar la forma de realizar la limpieza y carga de los datos	La explicación de esta actividad se la puede ver en la sección 3.1.4, en donde se explica cómo se procedió el análisis y limpieza de los datos.
2. Modelar los procesos ETL con BPMN	Anexo 2
3. Implementar los procesos ETL con la herramienta Kettle	Figura 20 de la sección 3.1.4.
4. Crear los cubos multidimensionales	El detalle de los cubos creados se lo puede revisar en la sección 3.1.4. (Creación de cubos multidimensionales)
5. Pruebas unitarias y de integración	Revisar en la sección 4.2.1.
6. Pruebas de validación de datos	Revisar en la sección 4.2.2.
7. Pruebas de rendimiento	Revisar en la sección 4.2.3.
8. Pruebas de regresión	Revisar en la sección 4.2.4.
9. Elaborar e implementar los dashboards y reportes para el usuario	Revisar en la sección 3.1.7

Fuente: Elaboración propia



Por último, en este trabajo se realizaron las pruebas de aceptación por parte de los usuarios. Cabe aclarar que el cuestionario se aplicó únicamente a dos personas, puesto que en la Comisión de Evaluación de la Universidad de Cuenca laboran únicamente el director y la secretaria. Cabe indicar también que se realizó una validación de forma personal con docentes y autoridades de Facultad, de manera que pudieron comprobar que los datos que se visualizan en el Datamart de Evaluación sí corresponden a las calificaciones finales obtenidas por los docentes. De esta manera se pudo concluir que el Datamart de Evaluación es de gran utilidad para la toma de decisiones, puesto que presenta valores confiables y oportunos.



Capítulo 4. VALIDACIÓN DE RESULTADOS

En este capítulo se describe el proceso de validación de los resultados obtenidos en este trabajo. Para lo cual se tomaron dos consideraciones, la primera consistió en verificar si el Datamart de evaluación permite contestar las preguntas identificadas en el componente de análisis (Figura 11) del capítulo tres. La segunda forma de realizar la validación fue aplicando pruebas de calidad de la solución implementada, haciendo énfasis en las pruebas de aceptación por parte de los usuarios finales con el objetivo de comprobar que se ha obtenido un software de calidad.

4.1 Verificación a través de las preguntas planteadas

A continuación, se procedió a verificar que el Datamart de evaluación pueda responder a las tres preguntas planteadas en la sección 3.1.1.

4.1.1 Primera pregunta: “Calificación obtenida por un docente en una facultad y carrera en un tiempo determinado”

Esta pregunta se contesta mediante el cubo llamado: 2014_evaluacion_docente, el mismo que fue elaborado para atacar a los procesos de evaluación realizados desde el año 2014 en adelante. Puesto que la evaluación en los primeros procesos no se obtenía por carrera, sino que llegaban al nivel de escuela. Las dimensiones tomadas en cuenta para este cubo son: Docente, Carrera, Edad, Facultad, Género y Tiempo. Para comprobar los resultados de esta pregunta, se puede ver un ejemplo en la Tabla 9, en donde se procedió a filtrar la información de algunos docentes por facultad, carrera y periodo de evaluación. En la Tabla se puede visualizar los procesos de evaluación ejecutados desde el año 2014. Se seleccionó la Facultad de Ingeniería y se escogieron algunos docentes para analizar las calificaciones que obtuvieron en cada una de las carreras pertenecientes a la Facultad. En el ejemplo se puede ver que en el proceso ejecutado en el año 2014 el docente “Víctor Saquicela” no tiene evaluación en la carrera de Ingeniería Civil a diferencia de los demás procesos que tiene calificación de la evaluación en las dos carreras que dictó clases en dichos periodos. Otro ejemplo



que se puede ver es el de la docente “Irene Cedillo”, que se puede observar que tiene calificación únicamente en el periodo 2016 y sólo para la carrera de Ingeniería de Sistemas.

Tabla 9: Calificación del docente por facultad, carrera y proceso de evaluación

Año o proceso de evaluación	2014		2015		2016	
	FACULTAD DE INGENIERÍA		FACULTAD DE INGENIERÍA		FACULTAD DE INGENIERÍA	
	INGENIERIA CIVIL	INGENIERIA DE SISTEMAS	INGENIERIA CIVIL	INGENIERIA DE SISTEMAS	INGENIERIA CIVIL	INGENIERIA DE SISTEMAS
Docente	Calificación		Calificación		Calificación	
SAQUICELA GALARZA VICTOR HUGO	-	92,77	95,11	95,11	94,01	94,01
SANCHEZ PERALTA MALHENA DE LOURDES	94,98	94,98	-	89,99	-	97,74
BERMEO CONTO JORGE LUIS	97,58	97,58	-	95,48	94,03	94,03
CEDILLO ORELLANA IRENE PRISCILA	-	-	-	-	-	98,87

Fuente: Elaboración propia

4.1.2 Segunda pregunta: “Calificación obtenida por el docente de parte de un actor en una función en un tiempo determinado”

Esta pregunta se pudo contestar para todos los procesos de evaluación; sin embargo, se separan los procesos por el hecho de que, a partir del 2014, a diferencia de los procesos anteriores, se incluye la calificación por carrera. Los cubos que contestan estas preguntas se denominan: 2013_evaluacion_funcion y 2014_evaluacion_funcion. En el primer caso intervienen las dimensiones sobre: docente, actor, edad, facultad, función, género y tiempo. En el segundo cubo además de estas dimensiones se incluye a la dimensión carreras.

En la Tabla 10 se puede visualizar los datos obtenidos para dar respuesta a la segunda pregunta, por ejemplo, para el caso de la docente “Rosa Ávila”, se puede visualizar las calificaciones que ha obtenido en cada función que le corresponde calificar a los actores, es por eso que se puede ver que los estudiantes solo califican la función docencia y los docentes y autoridades califican también las funciones de gestión e investigación. Sin embargo, se puede notar que la docente Ávila no es



calificada en la función investigación en ninguno de los procesos realizados, puesto que en su distributivo no tiene asignación de horas para esta actividad.

Tabla 10: Calificación de un docente por actor y función en un proceso de evaluación

Docente	Año o proceso de evaluación	AUTORIDADES			DOCENTES			ESTUDIANTES
		Función ADMINISTRACION / GESTION - DOCENCIA	Función DOCENCIA	Función INVESTIGACION	Función ADMINISTRACION / GESTION - DOCENCIA	Función DOCENCIA	Función INVESTIGACION	Función DOCENCIA
ROJAS REYES ROSENDO IVAN	2010_PROCESO EVALUAION 1	55	86,66	-	80	90,46	-	93,39
	2010_PROCESO EVALUAION 2	-	-	-	-	-	-	86,22
	2011	100	97,777	-	100	90,517	-	83,09
	2012	89,86	96,66	-	82,5	86,165	-	88,657
	2013	100	100	100	94,93	87,81	75,62	88,64
		-	-	-	-	-	-	-
TORRES UGALDE CARLOS ENRIQUE	2010_PROCESO EVALUAION 1	55	46,66	-	100	100	-	88,92
	2010_PROCESO EVALUAION 2	100	93,33	-	100	100	-	93,17
		-	-	-	-	-	-	-
AVILA GUARACA ROSA ISABEL	2010_PROCESO EVALUAION 1	90	93,33	-	81,667	93,227	-	84,08
	2010_PROCESO EVALUAION 2	96,667	94,443	-	100	99,06	-	82,68
	2011	92,395	91,66	-	93,333	99,06	-	85,013
	2012	94,93	91,665	-	100	98,59	-	88,545
	2013	95,015	93,33	-	100	96,555	-	88,185
		-	-	-	-	-	-	-
	Grand Total	-	-	-	-	-	-	-

Fuente: Elaboración propia

4.1.3 Tercera pregunta: “Número de participantes por facultad y género en un tiempo determinado”

La pregunta sobre el número de participantes en los procesos de evaluación se puede obtener en cualquiera de los cubos creados, sin embargo, para demostrar que el Datamart de evaluación puede responder a esta pregunta, se utilizó el cubo evaluación_docente_total, el resultado se puede visualizar en la Tabla 11, en la que se puede notar el número de participantes en los procesos de evaluación por cada uno de los actores participantes y clasificados según su género, por ejemplo se puede ver que en todos los procesos realizados, en la facultad de artes hay seis autoridades que evalúan, sin embargo, en el 2014 son tres hombres y tres mujeres, y en el 2015 y 2016 hay dos mujeres y cuatro varones. Otro ejemplo que se puede analizar es la participación de los estudiantes, se puede observar que el número de participantes tanto de hombres como de mujeres aumenta en todos los procesos en la Facultad de Arquitectura, a diferencia de la Facultad de Ciencias Económicas y Administrativas en donde se puede ver que entre el 2014 y el 2015 hay un incremento de participantes, no así entre el 2015 y 2016 donde hay una baja en el número de estudiantes que evalúan.



Tabla 11: Número de participantes por facultad y género en un proceso de evaluación

Facultad	Año o proceso de evaluación Actor	Género	2014		2015		2016	
			FEMENINO Participación	MASCULINO Participación	FEMENINO Participación	MASCULINO Participación	FEMENINO Participación	MASCULINO Participación
FACULTAD DE ARQUITECTURA Y URBANISMO	AUTORIDADES		0	3	0	3	0	3
	COMISION PARES ACADÉMICOS		1	1	1	1	2	3
	DOCENTES		16	48	22	44	22	42
	ESTUDIANTES		233	237	299	332	340	379
FACULTAD DE ARTES	AUTORIDADES		3	3	2	4	2	4
	COMISION PARES ACADÉMICOS		1	2	3	6	3	11
	DOCENTES		22	52	22	55	32	56
	ESTUDIANTES		345	333	498	490	509	517
FACULTAD DE CIENCIAS AGROPECUARIAS	AUTORIDADES		1	3	1	3	1	3
	COMISION PARES ACADÉMICOS		2	1	1	5	0	7
	DOCENTES		19	49	19	61	16	55
	ESTUDIANTES		298	316	460	440	511	477
FACULTAD DE CIENCIAS DE LA HOSPITALIDAD	AUTORIDADES		4	2	4	2	2	2
	COMISION PARES ACADÉMICOS		1	2	5	3	11	4
	DOCENTES		25	24	27	22	32	21
	ESTUDIANTES		399	249	500	334	590	354
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS	AUTORIDADES		4	6	2	7	2	6
	COMISION PARES ACADÉMICOS		0	3	9	12	6	8
	DOCENTES		44	68	46	72	41	75
	ESTUDIANTES		1388	538	1662	834	1477	808
FACULTAD DE CIENCIAS MÉDICAS	AUTORIDADES		8	2	10	1	2	2

Fuente: Elaboración propia

De lo expuesto anteriormente, se pudo comprobar que todas las preguntas planteadas para evaluación, son contestadas a través de los cubos construidos para el Datamart de evaluación.

4.2 Pruebas de calidad de la solución

Para realizar la verificación de la calidad del Datamart de evaluación se realizaron diferentes tipos de pruebas que se detallan a continuación:

4.2.1 Pruebas unitarias y de integración

Es importante realizar pruebas unitarias sobre los procesos ETL, con el objetivo de garantizar que los datos sean consistentes y también para comprobar que los ETL cumplan con su propósito. Las pruebas unitarias se realizaron como parte del componente de validación de resultados. Luego de realizar las pruebas unitarias se ejecutaron las pruebas de integración, las que consistieron en verificar que se superaron los inconvenientes que se presentaban durante la ejecución de la aplicación para lo cual se probaron cada uno de los cubos filtrando la información de varias formas y de manera combinada.



4.2.2 Pruebas de validación de datos

Estas pruebas se realizaron, en base a una comparación entre la información que se obtiene a través de varios reportes y consultas que actualmente están implementadas en el sistema de gestión de evaluación con la información consultada a través de la herramienta Saiku Analytics de Pentaho (Saiku Analytics 3.x documentation, s.f.).

Cabe mencionar que actualmente el sistema de evaluación posee varios reportes y que para extraer información de las preguntas planteadas en este trabajo se tiene que ejecutar varias consultas a través de estos reportes, luego exportar a Excel y combinarlos. Al inicio de este tipo de pruebas se pudo encontrar varias inconsistencias en los valores proporcionados por el Datamart, los que fueron corregidos al revisar los ETL y la forma de cálculo de las calificaciones de los docentes en la tabla de hechos, esto tanto para los procesos iniciales es decir los realizados a partir del 2010 hasta el 2013 y también los realizados a partir del año 2014 hasta la fecha.

4.2.3 Pruebas de rendimiento

Las pruebas de rendimiento se realizaron conjuntamente con las pruebas de validación de datos, puesto que al momento que se comparaban los resultados del Datamart con los de los sistemas informáticos actuales, se podía evidenciar problemas en el aplicativo, por lo que fue necesario revisar las configuraciones tanto del servidor de aplicaciones como de la base de datos y realizar los ajustes necesarios para solucionar este tipo de inconvenientes y obtener de esta manera un software de calidad a la medida del usuario.

Para probar el rendimiento de la aplicación se lo hizo midiendo los tiempos de página y transacciones, en donde se pudo observar que al inicio la respuesta del aplicativo al cargar las páginas con los resultados a consultas realizadas no se daba en tiempos aceptables por lo que se procedió a corregir las configuraciones tanto en el servidor como en la base de datos.



4.2.4 Pruebas de regresión

Este tipo de pruebas se aplicó en este trabajo luego de cada revisión con el usuario final, para comprobar que se sienta satisfecho al utilizar el Datamart de evaluación y obtener resultados de manera más rápida y efectiva, evitando el análisis de la información de los reportes del SGE de manera separada para llegar al resultado deseado.

4.2.5 Pruebas de aceptación de usuario

Estas pruebas se realizaron conjuntamente con los usuarios, a través de la revisión y validación de las funcionalidades de la solución implementada, con el propósito de comprobar que cumple con las expectativas de los usuarios finales, en este caso se trata de los funcionarios y técnicos que laboran en los procesos de evaluación de la Universidad. Luego de realizadas las pruebas, estas arrojaron los siguientes resultados.

Una prueba más que se ejecutó para medir la calidad del software desde el punto de vista del usuario final fue aplicar un cuestionario de 10 preguntas, el mismo que se basa en el cuestionario SUMI (Software Usability Measurement Inventory) y cuyas respuestas se debe contestar mediante tres opciones “De acuerdo”, “Indeciso” o “En desacuerdo” (SUMI, s.f.). A continuación, en la Tabla 12 se visualiza el cuestionario que fue respondido por un usuario final para evaluar la implementación del Datamart de Evaluación. El cuestionario fue aplicado a dos usuarios que laboran en el área de evaluación de la Institución.

Tabla 12: Cuestionario aplicado a usuario final

Pregunta	Opciones	Opciones		
		De acuerdo	Indeciso	En desacuerdo
1	El Software responde demasiado lento a sus entradas	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
2	Recomendaría este software a otros usuarios	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Las instrucciones e indicaciones sirven de ayuda	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
4	El software en algún momento se detuvo inesperadamente	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>



5	Al iniciar el uso del software es difícil aprender a utilizarlo	⊙	⊙	⊙
6	A veces no sé qué hacer con este software	⊙	⊙	⊙
7	Disfruto el tiempo que utilizo este software	⊙	⊙	⊙
8	La información de ayuda dada por este software no es muy útil	⊙	⊙	⊙
9	Si el software se detiene, no es fácil reiniciarlo	⊙	⊙	⊙
10	Es demorado aprender a utilizar las funcionalidades del software	⊙	⊙	⊙

Fuente: Elaboración propia basado en (SUMI, s.f.)

En la Tabla 12, se puede ver las respuestas a cada una de las preguntas del cuestionario realizadas a uno de los usuarios. Luego de aplicar el cuestionario se clasificaron las respuestas en positivas, negativas e indecisas. Estos resultados se pueden visualizar en la Tabla 13, y en la Tabla 14 se puede observar un resumen del total de las respuestas según como fueron clasificadas.

Tabla 13: Contabilización de respuestas al cuestionario

Pregunta	Respuesta		
	Positiva	Negativa	Indecisa
El Software responde demasiado lento a sus entradas			1
Recomendaría este software a otros usuarios	1		
Las instrucciones e indicaciones sirven de ayuda			1
El software en algún momento se detuvo inesperadamente	1		
Al iniciar el uso del software es difícil aprender a utilizarlo	1		
A veces no sé qué hacer con este software			1
Disfruto el tiempo que utilizo este software			1
La información de ayuda dada por este software no es muy útil	1		
Si el software se detiene, no es fácil reiniciarlo	1		
Es demorado aprender a utilizar las funcionalidades del software	1		

Fuente: Elaboración propia

Tabla 14: Resumen de las respuestas de manera clasificada

Tipo de respuesta	No. De Respuestas	Porcentaje
Respuestas Positivas	6	60%
Respuestas negativas	0	0%
Respuestas Indecisas	4	40%

Fuente: Elaboración propia



En la Tabla 14 se puede observar que existen 6 respuestas positivas y que no existen respuestas negativas, sin embargo: las 4 respuestas indecisas se atribuyen a que es un nuevo aplicativo y que estas respuestas pueden cambiar con el tiempo, luego de que el usuario se familiarice más con el aplicativo. De lo anteriormente expuesto se puede concluir que existe un 60 % de aceptación del aplicativo por parte del usuario que participó como Product Owner (dueño del producto). Como se puede observar en la Tabla 14, a través de esta prueba también se obtienen resultados favorables por parte del usuario final.

Las pruebas de aceptación permiten concluir que el acceso a la información a través de la solución planteada desde el punto de vista del usuario es más fácil puesto que las herramientas de análisis de información de cubos de datos son intuitivas y permiten la manipulación de las dimensiones y hechos de manera fácil.



Capítulo 5. CONCLUSIONES Y TRABAJOS FUTUROS

Es importante indicar que un DataWarehouse para una empresa no se compra, sino que se lo construye y 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, eficaz y desde diferentes puntos de vista.

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. Para lograr el éxito deseado en la implementación de la solución planteada, se trabajó de manera estrecha con los usuarios de tal forma que comprendieran los beneficios de esta implementación. Con la colaboración cercana y activa de los usuarios, se logró corregir incidentes y errores a tiempo. 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.

Al realizar la validación del aplicativo con los usuarios en este trabajo, se pudo comprobar que, para el Departamento de Evaluación de la Universidad, el DataWarehouse es la herramienta que permite obtener la información útil para realizar los análisis correspondientes a las calificaciones que los docentes han obtenido en cada proceso de evaluación, y en base a ésta tomar importantes decisiones. La información que se obtiene del DataWarehouse permitió que los funcionarios del área de evaluación de la Institución ya no realicen procesos manuales y no dependan de los técnicos del área de tecnología para conseguir la información requerida.

Durante la implementación del Datamart de Evaluación se pudo comprobar que éste aporta significativamente la consecución de los objetivos planteados en el Plan Estratégico 2012 - 2017 de la Universidad de Cuenca, mencionados en el componente de análisis, sección 1.3.1. Se puede ver



que estos objetivos estratégicos se pueden dar respuesta a través del análisis de información y la extracción de indicadores, por lo que se concluye que al implementar todo el DataWarehouse de la Institución, éste podrá aportar de mejor manera a la consecución de los objetivos estratégicos de cada uno de sus ejes misionales a través del análisis de información.

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.

Por último, es importante recomendar que esta herramienta sea difundida y de uso generalizado por parte de las autoridades de la Institución, puesto que, así como se dio resultados favorables en el área de evaluación al docente, también será de utilidad para el resto de las áreas estudiadas y analizadas al inicio de este trabajo. A fecha de hoy se tienen en proceso de implementación los Datamarts de talento humano y el de proyectos, los cuales siguen el procedimiento planteado en este trabajo. Al terminar la implementación de estos dos Datamarts se continuará con los demás Datamarts que fueron definidos en la Tabla 1 del componente de análisis.



Glosario de términos

SGI:	Sistema de Gestión de Investigación
SGA:	Sistema de Gestión Académica
SGE:	Sistema de Gestión de Evaluación al desempeño del docente
SGP:	Sistema de Gestión de Personal
AdminUC:	Base de datos común para todas las aplicaciones de la Universidad, que contiene información de las personas, usuarios, etc..
DataWarehouse:	Almacén de datos, que contiene toda la información centralizada de la empresa
Datamart:	Conjunto de datos que pertenecen a un área específica de la empresa
ETL:	(Extract, Transform and Load) Proceso para extraer la información de los sistemas operacionales, transformarlos y limpiarlos y luego realizar la carga en el DataWarehouse
Redmine:	Herramienta para la gestión de proyectos
SGAP:	Sistema de Gestión Académica de Posgrados
BPMN:	(Business Process Model and Notation) Notación gráfica estandarizada que se utiliza para modelar los procesos de un negocio
SUMI:	(Software Usability Measurement Inventory) Cuestionario para medir la calidad del software
IES:	Institución de Educación Superior



Referencias bibliográficas

- Alanocca, Y., & Ccahuana, J. (2015). *Sistema ERP aplicando SCRUM para optimizar la administración e integración de información entre las diferentes áreas de las empresas retail del Perú - 2013*. Tesis de grado, Universidad Nacional del Altiplano, Facultad de Ingeniería Mecánica Eléctrica, Electrónica y Sistemas, Puno, Perú.
- Bernabeu, R. (2010). *HEFESTO*. Córdoba, Argentina.
- Castellano, R. (2000). *wikispaces.com*. Obtenido de <https://ti-1.wikispaces.com/5.1+Sistemas+de+Negocios>
- Castillo, I. (2014). *Diseño, elaboración y explotación de un data warehouse para una institución sanitaria*. Proyecto de fin de carrera.
- CEAACES. (s.f.). Obtenido de Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior: <http://ceaaces.gob.ec>
- CES. (s.f.). Obtenido de Consejo de Educación Superior: <http://www.ces.gob.ec>
- Chinchilla, R. (2011). Mercado de datos: conceptos y metodologías de desarrollo. *Revista Tecnología en Marcha*, 55-56.
- Conn, S. (2005). OLTP and OLAP data integration: a review of feasible implementation methods and architectures for real time data analysis. *SoutheastCon, 2005. Proceedings. IEEE*, 515-520.
- Curto, J. (2012). *Introducción al Business Intelligence*. Barcelona: Editorial UOC.
- Factos, C. (2014). *Integración y consumo de información del data warehouse del Sistema Financiero Ecuatoriano, diseño, análisis de reportes dinámicos y su visualización web y en dispositivos móviles*. Tesis de grado, Universidad Central del Ecuador, Facultad de Ingeniería, Ciencias Físicas y Matemática, Quito, Ecuador.
- Figueroa, R., Solis, C., & Cabrera, A. (2008). *Adonisnet's Weblog*. Obtenido de <http://adonisnet.files.wordpress.com/2008/06/articulo-metodologia-de-sw-formato.doc>
- Gómez, J. (2014). Una Mirada a BPMN como Herramienta Estándar para el Modelado de Procesos de Negocio. *CIIS 2014*, (págs. 130-142).
- Guan, J., Nunez, W., & Welsh, J. F. (2002). Institutional strategy and information support: the rol of data warehousing in higher education. *Campus-Wide Information Systems*, 19(Iss 5), 168-174.
- Inmon, W. H. (1996). The data warehouse and data mining. *Association for Computing Machinery. Communications of the ACM*, 49+.
- Inmon, W. H. (2002). *Building the Data Warehouse, 3rd Edition*. New York: John Wiley & Sons, Inc.
- Kimball, R. (1996). *The Data Warehouse Toolkit*. New York, NY: John Wiley & Sons.



- Ley Orgánica de Educación Superior, LOES.* (2010). Obtenido de <http://www.conocimiento.gob.ec/wp-content/uploads/2015/07/Ley-Organica-de-Educacion-Superior-LOES.pdf>
- Lozada, X., & Cruz, H. (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.* Tesis de grado, Universidad de las Fuerzas Armadas .
- Menéndez, M., & De Luján, M. (2012). Sistemas para la toma de decisiones en el ámbito universitario. *6º Simposio Argentino De Informática En El Estado*, (págs. 164-175). Buenos Aires, Argentina.
- Mulder, S. (2010). *An Action Research study on the use of Scrum to provide agility in Data Warehouse development.* Requisito de grado, University of Pretoria, M. Com. in Informatics.
- Naranjo, E., & Sáenz, I. (2012). Pentaho: software líder de Inteligencia de Negocio de código abierto. *Revista Telemática*, 10(2). Obtenido de <http://revistatelematica.cujae.edu.cu/index.php/tele/article/view/44>
- Pirani, J., & Albrecht, B. (2005). University of Phoenix: Driving decisions through academic analytics. *EDUCAUSE Center for Applied Research*.
- Reimundo, A., Nasimba, V., & Cueva, P. (s.f.). *Metodología SCRUM en proyectos Business Intelligence.* Obtenido de Prezi: <https://prezi.com/fad-el4f6e1r/metodologia-scrum-en-proyectos-bissines-intelligence/>
- Rivadera, G. (2010). La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). *Universidad Católica de Salta, Cuadernos de la Facultad*, 56-71.
- Saiku Analytics 3.x documentation.* (s.f.). Obtenido de <http://saiku-documentation.readthedocs.io/en/latest/>
- Sanchez, R. (26 de Octubre de 2003). *gestiopolis.* Obtenido de <https://www.gestiopolis.com/inteligencia-de-negocios-o-business-intelligence-bi/>
- Scrum.org.* (s.f.). Obtenido de <https://www.scrum.org/>
- SUMI.* (s.f.). Obtenido de <http://sumi.uxp.ie/>
- Universidad de Cuenca.* (s.f.). Obtenido de <http://ucuenca.edu.ec>
- White, S., & Miers, D. (2010). *BPMN Guía de Referencia y Modelado Comprendiendo y Utilizando BPMN.*



ANEXO 1. Identificación de la Empresa - UNIVERSIDAD DE CUENCA

La Universidad de Cuenca es una universidad pública, tiene su sede en la capital de la provincia del Azuay a su vez cabecera de la región centro-sur del Ecuador. Creada por decreto legislativo del 15 de octubre de 1867, la Universidad de Cuenca se denominó inicialmente Corporación Universitaria del Azuay (Universidad de Cuenca, s.f.)

Misión

“La Universidad de Cuenca es una universidad pública, cuya misión es formar profesionales y científicos comprometidos con el mejoramiento de la calidad de vida, en el contexto de la interculturalidad y en armonía con la naturaleza. La Universidad fundamenta en la calidad académica, en la creatividad y en la innovación, su capacidad para responder a los retos científicos y humanos de la época y sociedad regional, nacional e internacional equitativa, solidaria y eficiente.”

Visión

“La Universidad de Cuenca se proyecta como una institución con reconocimiento nacional e internacional por su excelencia en docencia con investigación y vinculación con la colectividad; comprometida con los planes de desarrollo regional y nacional; que impulsa y lidera un modelo de pensamiento crítico en la sociedad.”

Políticas Institucionales

1. Fortalecer el desarrollo de la ciencia, la tecnología y la innovación, con enfoque disciplinario e interdisciplinario, promoviendo la inclusión de conocimientos y saberes ancestrales, y en respuesta a las necesidades sociales.
2. Mejorar progresivamente la calidad de la educación superior con visión científica y humanista, fortaleciendo el modelo educativo centrado en el aprendizaje del estudiante, con la aplicación de



estándares internacionales de calidad, y articulada a los objetivos del Plan Nacional para el Buen Vivir.

3. Fortalecer los vínculos institucionales con organizaciones públicas, privadas y comunitarias, identificando campos de acción prioritarios para la colaboración y cooperación mutua.
4. Impulsar procesos de mejora continua en la gestión institucional como actividades de soporte a la gestión académica, operativa y de direccionamiento estratégico.

Valores

- Compromiso
 - Servir a la sociedad y promover la preservación del medio ambiente.
 - Cumplir con las regulaciones legales y reglamentarias.
 - Apoyar al cumplimiento de las metas del PNBV en lo pertinente a las IES.
 - Fortalecer el sentido de identidad y pertenencia aportando proactivamente a las estrategias de mejoramiento institucional
- Transparencia
 - Transparentar todos los actos académicos, científicos y administrativos
 - Facilitar el acceso del público a la información institucional.
 - Presentar informes y rendir cuentas a la comunidad universitaria y a la sociedad
- Excelencia
 - Fomentar e impulsar cambios en la calidad y pertinencia de la educación superior.
 - Trabajar bajo los principios de calidad y pertinencia social y científica en el cumplimiento de los ejes misionales
 - Gestionar la excelencia con eficacia y eficiencia
 - Liderar la gestión académica, científica y administrativa.
- Lealtad
 - Cumplir con la visión, misión y objetivos institucionales aportando a la consecución de las metas del Plan Nacional para el Buen Vivir y al Plan Nacional de Ciencia, Tecnología e Innovación.
 - Fomentar el trabajo en Equipo.
 - Asegurar una comunicación altamente efectiva hacia dentro y hacia fuera.

Apoyar a la gestión institucional.
- Innovación
 - Generar nuevo conocimiento científico y tecnológico.
 - Gestionar los cambios institucionales y del entorno con visión proactiva.
 - Fortalecer las capacidades y competencias del talento humano.
 - Impulsar el desarrollo tecnológico para mejorar la gestión académica y de investigación.



- Equidad
 - Generar en la comunidad universitaria hábitos de autorreflexión organizacional para promover el cambio.
 - Mejorar y diversificar las políticas de acción afirmativa.

Estrategias y líneas de acción

Las estrategias se describen por cada eje misional

- En Ciencia, Tecnología e Innovación
Creación ordenada de departamentos y centros de investigación en las áreas científico técnica, biológicas y de la salud y ciencias sociales, humanidades y artes. Se cuenta ya para ello con un Reglamento de Departamentos de la Universidad de Cuenca aprobado.

Líneas de Acción. -

- Redefiniendo la estructura organizacional para investigación que permita la implementación de departamentos de investigación y la escuela internacional de posgrado.
 - Creando ordenadamente departamentos y centros de investigación en las áreas científico técnica, biológicas y de la salud y ciencias sociales, humanidades y artes. Se cuenta con el Reglamento de Departamentos de la Universidad de Cuenca aprobado por el H Consejo Universitario.
 - Promoviendo la conformación de equipos de investigación disciplinarios e interdisciplinarios.
 - Incentivando al personal de investigación a mejorar sus competencias.
 - Apoyando la inserción de los equipos de investigación en redes de investigación y posgrado a nivel nacional e internacional.
 - Promoviendo investigaciones e innovaciones de calidad en alianza con los sectores gubernamental, académico y productivo.
 - Incorporando e institucionalizando la investigación formativa en las mallas curriculares.
- En Docencia
Revisión coordinada de mallas curriculares de grado desde un enfoque de docencia con investigación y la institucionalización e internacionalización del posgrado (Escuela Internacional de Posgrado), en coherencia con el nuevo Reglamento Académico.

Líneas de Acción. -

- Garantizando la excelencia académica con políticas para una universidad del siglo XXI.



- Promoviendo la formación de cuarto nivel del personal académico (M.Sc, PhD y pos PhD).
 - Revisando coordinadamente las mallas curriculares de grado desde un enfoque de docencia con investigación y la institucionalización e internacionalización del posgrado (Escuela Internacional de Posgrado).
 - Implementado un modelo de educación centrado en el aprendizaje.
 - Acompañando al estudiante en su proceso de formación académica y su posterior inserción en IES internacionales para formación de cuarto nivel, dotándole de las competencias y herramientas para su autoaprendizaje.
- En Vinculación con la colectividad
Buscar una vinculación integral a nivel zonal y nacional en todas sus vertientes: matriz productiva, sociedad, universidades tradicionales y de nueva creación. A nivel internacional, integración en redes de colaboración y redes de excelencia de docencia e investigación

Líneas de Acción. -

- Buscando la vinculación integral a nivel zonal y nacional en todas sus vertientes: matriz productiva, sociedad, universidades tradicionales y de nueva creación. A nivel internacional, integración en redes de colaboración y redes de excelencia de docencia e investigación.
 - Fomentando la autogestión de la UC mediante la transferencia de resultados de investigación a la colectividad, brindando servicios y consultorías de calidad.
 - Analizando el impacto de la formación de los graduados de la institución en la sociedad.
- En Gestión Institucional
Redefinir considerablemente las tareas administrativas minimizando el trayecto hacia ellas desde las diferentes unidades académicas a través de una nueva estructura organizacional derivada del nuevo modelo de universidad.

Líneas de Acción. -

- Agilizar considerablemente las tareas administrativas minimizando el trayecto hacia ellas desde las diferentes unidades académicas a través de una nueva estructura organizacional derivada del nuevo modelo de universidad.
- Incorporando un modelo de gestión institucional basado en procesos
- Formando al talento humano administrativo en base a sus competencias y necesidades institucionales.
- Innovando la infraestructura física, tecnológica, bibliográfica y de equipamiento para responder a las necesidades de crecimiento y desarrollo institucional.
- Aplicando un plan de gestión integral de riesgos académicos y administrativos que permita mitigarlos, a fin de alcanzar con éxito los objetivos institucionales.
- Implementar un proceso de mejora continua en la prestación de los servicios informáticos.



Entorno

Hoy en día, la calidad de la educación superior es una de las prioridades del gobierno ecuatoriano, es por esta razón que en el Ecuador se promueve la eficiencia y la calidad de las Universidades y sus carreras. Todas las Instituciones de Educación Superior (IES) deben demostrar que siguen los criterios y normas de calidad establecidos por los distintos órganos de gobierno, estos son academia, investigación, gestión, entre otros. Para cumplir con estos criterios y normas las IESs deben contar con la información necesaria para realizar una autoevaluación y posterior acreditación por parte de los organismos competentes. La Universidad de Cuenca, día a día opera sus sistemas informáticos, obteniendo una gran cantidad de datos que pueden ser utilizados para el fin antes nombrado. Por esta razón es necesario que la Institución cuente con una herramienta que sea de ayuda para las autoridades en la toma de decisiones. La solución planteada en este trabajo consiste en implementar un *DataWarehouse* para la Universidad que consolide los datos y que permita obtener información de manera confiable, rápida y eficiente.

Estructura Organizacional de la Universidad de Cuenca

El organigrama organizacional de la Universidad de Cuenca, vigente según el estatuto de la Universidad, se visualiza gráficamente en la Figura 24

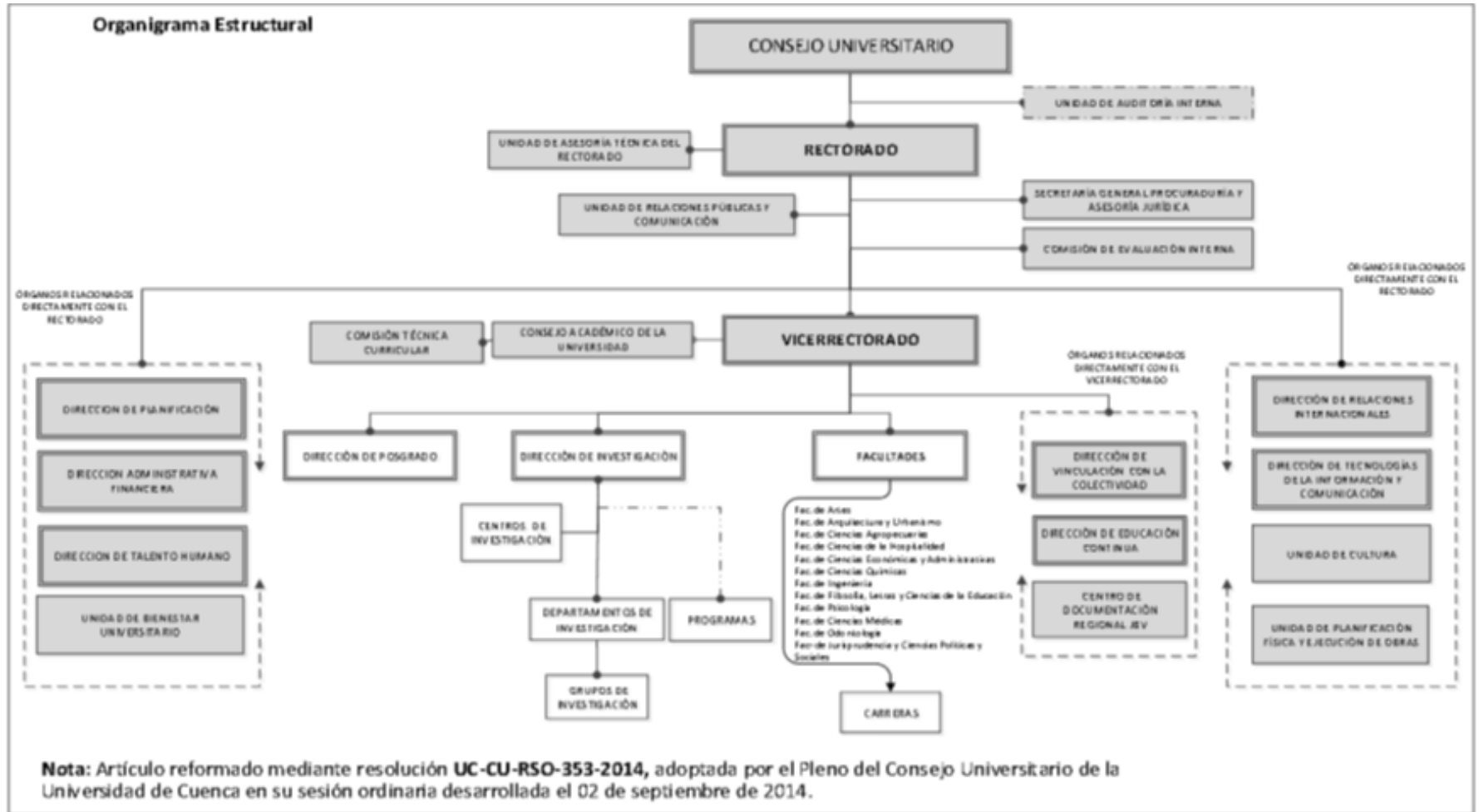


Figura 24: Organigrama de la Universidad de Cuenca

Fuente: Organigrama Estructural. [Figura]. Recuperado de <https://ucuenca.edu.ec>



Relación de los objetivos estratégicos con los objetivos del DataWarehouse

El plan estratégico de la Universidad de Cuenca plantea el objetivo principal que consiste en convertirse en una Universidad de docencia con investigación y se ha planteado también como objetivo llegar a ser una Institución de Educación Superior de excelencia a nivel nacional e internacional, para ello debe cumplir los diferentes aspectos considerados en el marco de referencia establecido por el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES), los criterios de evaluación que debe cumplir son: academia, eficiencia académica, investigación, organización e infraestructura. A través de la implementación del DataWarehouse la Universidad tendrá una herramienta que permita obtener fácilmente la información de los indicadores de forma rápida y confiable para realizar una evaluación del estado de la Institución y verificar que ha alcanzado sus metas.

A continuación, se nombran algunas ventajas de la implementación del DataWarehouse en la Universidad de Cuenca:

- Transforma los datos recolectados diariamente a través de los sistemas operacionales de la Universidad en información que sirve para realizar análisis y ayuda a la toma de decisiones.
- Proporciona reportes gráficos y dinámicos, útiles para el análisis.
- Estará alineado a los objetivos estratégicos de la Universidad, puesto que permitirá obtener los indicadores establecidos por el CEAACES.

Áreas de aplicación del DataWarehouse

La Universidad necesita realizar análisis de algunas áreas que son:

- Información de carreras y programas de posgrado.
- Información de evaluación del desempeño del docente.
- Información de talento humano.
- Información relativa a la investigación.

ANEXO 2. Procesos ETL

En este anexo se podrán encontrar como se realizó el modelamiento de cada uno de los procesos ETL utilizando la notación BPMN.

Modelamiento BPMN

Carga de la dimensión: dim_Tiempo

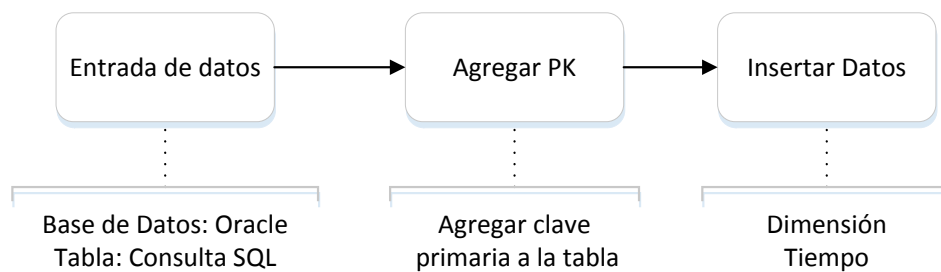


Figura 25: Carga para la dimensión tiempo (procesos de evaluación)

Fuente: Elaboración propia

Carga de la dimensión: dim_Docente

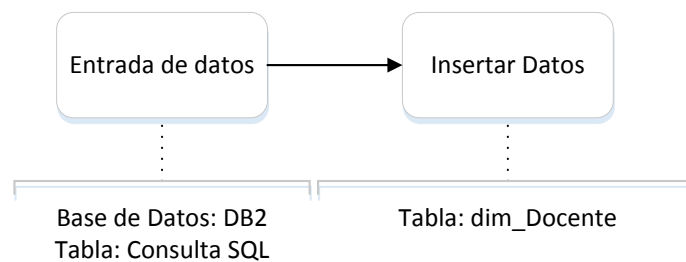


Figura 26: Carga para la dimensión docente

Fuente: Elaboración propia

Carga de las dimensiones: dim_Facultad y dim_Escuela

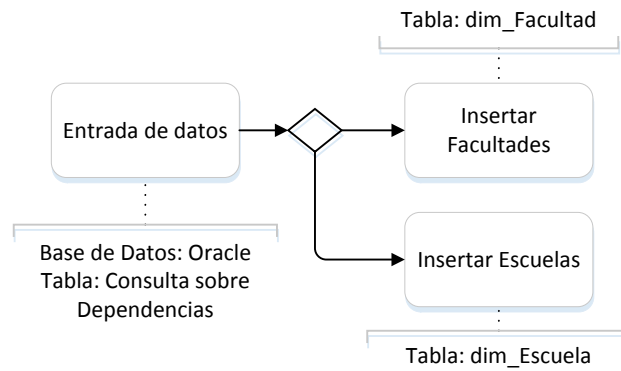


Figura 27: Carga para las dimensiones facultad y escuela

Fuente: Elaboración propia

Carga de las dimension: dim_Actor

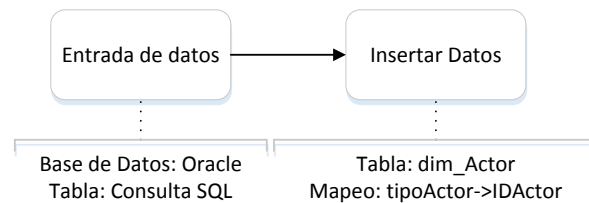


Figura 28: Carga para la dimensión actor

Fuente: Elaboración propia

Carga de las dimension: dim_Edad

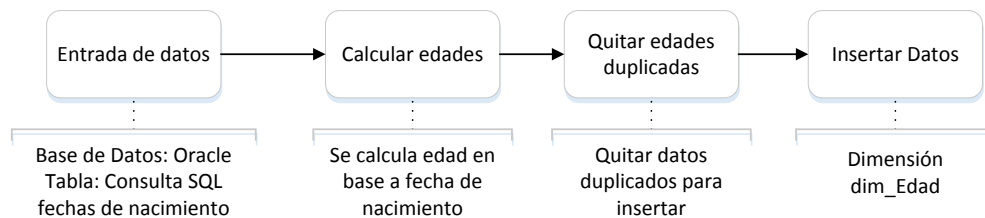


Figura 29: Carga para la dimensión edad

Fuente: Elaboración propia

Carga de las dimension: dim_Genero

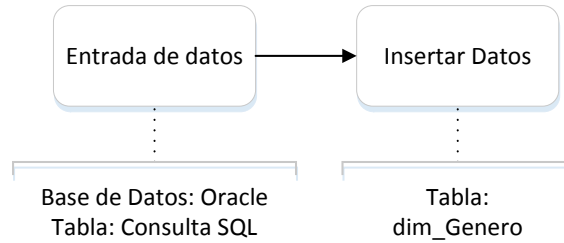


Figura 30: Carga para la dimensión genero

Fuente: Elaboración propia

Limpieza y carga de las evaluaciones docentes

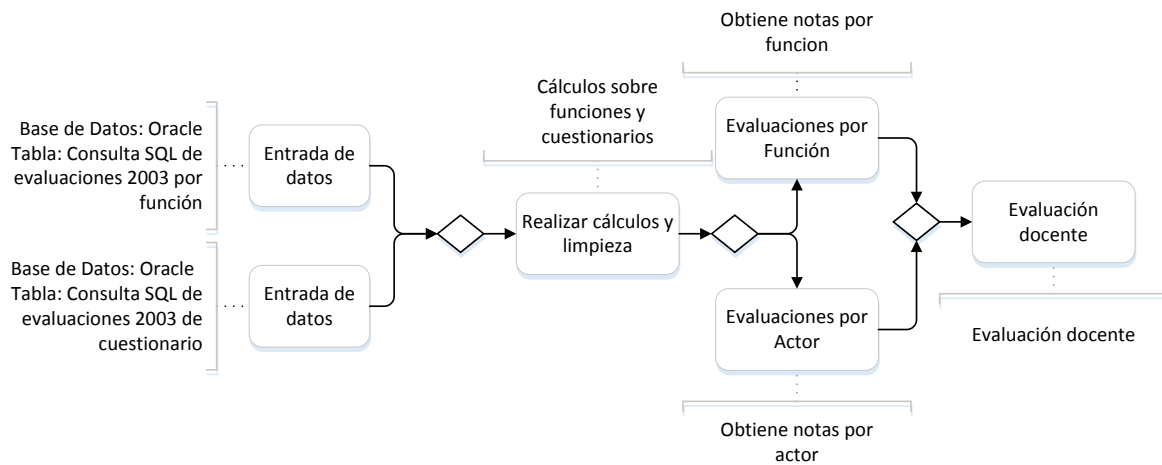


Figura 31: Limpieza y carga de datos para las evaluaciones

Fuente: Elaboración propia

Procesos ETL en herramienta Pentaho

ETL para las Dimensiones

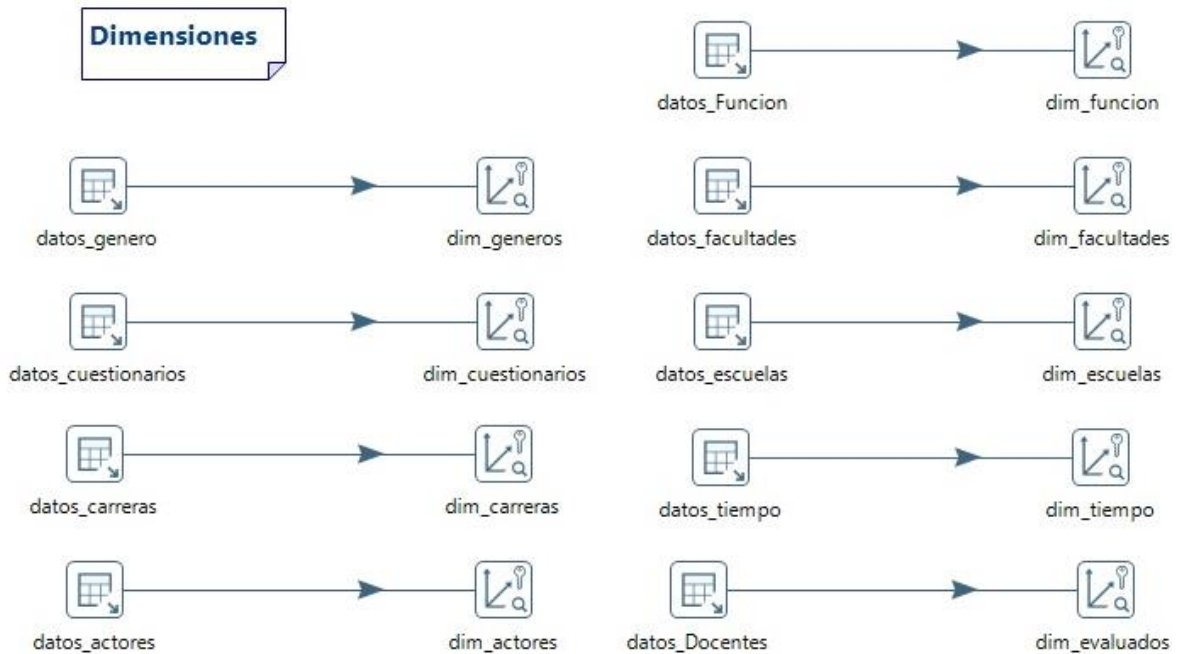


Figura 32: ETL para dimensiones

Fuente: Elaboración propia

ETL para Tabla de hechos



Figura 33: ETL para tabla de hechos

Fuente: Elaboración propia



ANEXO 3. Base de datos para el Sistema de Gestión de Evaluación del docente

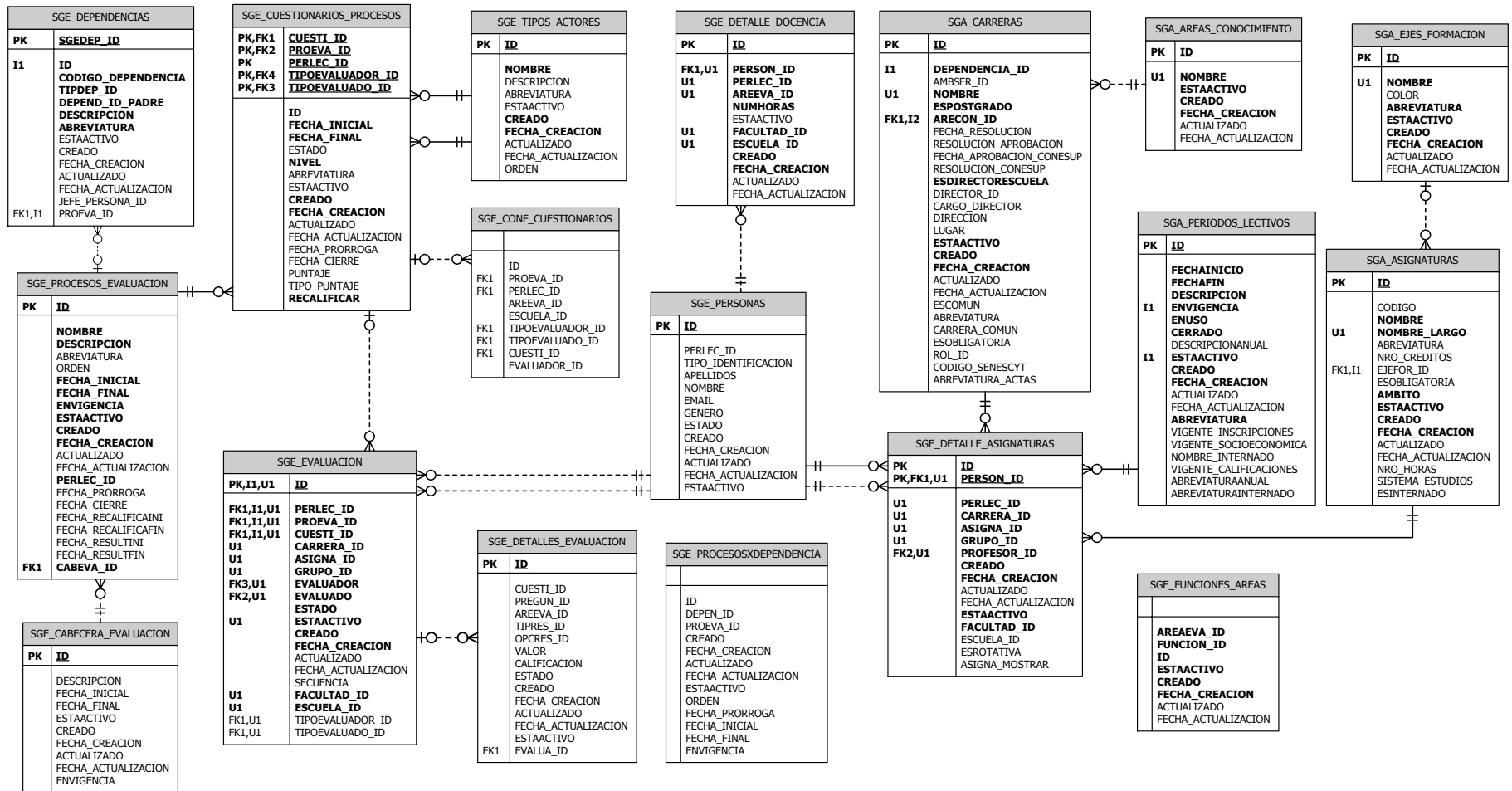


Figura 34: Modelo Entidad – Relación SGE

Fuente: Elaboración propia



Diccionario de Datos

Tabla 15: Diccionario de datos del esquema entidad-relacion del SGE

Tabla: SGE_AREAS_EVALUACION: Registros de áreas de evaluación				
PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador del área de evaluación
	DESCRIPCION	Varchar2(255)	N	Nombre del área de evaluación
	AREEVA_PADRE_ID	Number(22)	Y	Identificador del área padre.
	TIPARE_ID	Number(22)	N	Tipo de área de evaluación
	NIVEL	Number(22)	Y	Nivel jerárquico del registro.
	ESTAACTIVO	Char(1)	N	Borrado lógico del registro
	CREADO	Varchar2(20)	N	Usuario que crea el registro.
	FECHA_CREACION	Date(7)	N	Fecha de creación.
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza los datos.
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización de los datos
	ABREVIATURA	Varchar2(15)	Y	Abreviatura del área de evaluación.
	ORDEN	Number(22)	Y	Criterio de ordenamiento
	ESOPCIONAL	Varchar2(1)	N	Indica si el área de evaluación es opcional. Es decir las preguntas se responde dependiendo de ciertos criterios.
	ESVALORADO	Varchar2(1)	N	Las preguntas que se encuentran dentro de un área con este campo en valor 's', es decir valorado, entran al cómputo del resultado final de cuestionario

Tabla: SGE_CABECERA_EVALUACION: Evaluaciones generales				
PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de cabecera evaluación
	DESCRIPCION	Varchar2(4000)	Y	Descripción de cabecera evaluación
	FECHA_INICIAL	Date(7)	Y	Fecha de inicio de vigencia
	FECHA_FINAL	Date(7)	Y	Fecha de final de vigencia
	ESTAACTIVO	Char(1)	Y	S: Registro activo; n: registro inactivo.
	CREADO	Varchar2(15)	Y	Identificación de creación
	FECHA_CREACION	Date(7)	Y	Fecha creación
	ACTUALIZADO	Varchar2(15)	Y	Identificación actualización
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha actualización
	ENVIGENCIA	Varchar2(1)	Y	S: está vigente; n: no está vigente

Tabla: SGE_CONF_CUESTIONARIOS: Configuraciones de cuestionarios				
PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
	ID	Number(22)	Y	Id de la tabla



PROEVA_ID	Number(22)	Y	Id del proceso de evaluación
PERLEC_ID	Number(22)	Y	Id del periodo lectivo
AREEVA_ID	Number(22)	Y	Id del área de evaluación
ESCUELA_ID	Number(22)	Y	Id de la escuela
TIPOEVALUADOR_ID	Number(22)	Y	Id del tipo de evaluador
TIPOEVALUADO_ID	Number(22)	Y	Id del tipo evaluado
CUESTI_ID	Number(22)	Y	Id del cuestionario
EVALUADOR_ID	Varchar2(15)	Y	Id del evaluador

Tabla: SGE_CUESTIONARIOS: Almacena los cuestionarios

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de la tabla cuestionarios
	TIPCUE_ID	Number(22)	N	Tipo de cuestionario
	NOMBRE	Varchar2(50)	N	Nombre del cuestionario
	DESCRIPCION	Varchar2(255)	Y	Descripción del cuestionario
	INSTRUCCIONES	Varchar2(4000)	Y	Instrucciones para llenar cuestionario
	ABREVIATURA	Varchar2(15)	Y	Abreviatura del cuestionario
	TOOLTIP	Varchar2(255)	Y	Tooltip que describe el cuestionario
	ORDEN	Number(22)	Y	Criterio de ordenamiento
	ESTAACTIVO	Char(1)	N	Borrado lógico del registro
	DESCRIPCION1	Varchar2(4000)	Y	Descripción adicional de los cuestionarios
	DESCRIPCION2	Varchar2(4000)	Y	Descripción adicional de los cuestionarios
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro

Tabla: SGE_CUESTIONARIOS_PREGUNTAS: Almacena relación de cuestionarios con preguntas

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	CUESTI_ID	Number(22)	N	Identificador que indica hace referencia un cuestionario
S	PREGUN_ID	Number(22)	N	Identificador de la pregunta
S	AREEVA_ID	Number(22)	N	Identificador del área de evaluación
	ORDEN	Number(22)	Y	Campo que sirve para ordenar los registros dentro de una lista
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	ID	Number(22)	N	Identificador único de la tabla.

Tabla: SGE_CUESTIONARIOS_PREGUNTAS: Almacena relación de cuestionarios con preguntas



PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
	ID	Number(22)	N	Identificador único de la tabla.
S	CUESTI_ID	Number(22)	N	Identificador del cuestionario
S	PROEVA_ID	Number(22)	N	Identificador del proceso de evaluación
S	PERLEC_ID	Number(22)	N	Identificador del periodo lectivo
S	TIPOEVALUADOR_ID	Number(22)	N	Identificador del tipo de evaluador
S	TIPOEVALUADO_ID	Number(22)	N	Identificador del tipo de evaluado
	FECHA_INICIAL	Date(7)	N	Fecha en la que se inicia la evaluación
	FECHA_FINAL	Date(7)	N	Fecha de finalización de la evaluación
	ESTADO	Char(1)	Y	Estado
	NIVEL	Varchar2(255)	N	Nivel
	ABREVIATURA	Varchar2(15)	Y	Abreviatura del cuestionario proceso
	ESTAACTIVO	Char(1)	Y	Borrado lógico del registro
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	FECHA_PRORROGA	Date(7)	Y	Fecha de prórroga del proceso.
	FECHA_CIERRE	Date(7)	Y	Fecha de cierre de la evaluación
	PUNTAJE	Number(22)	Y	Puntaje que tiene el cuestionario dentro de un proceso de evaluación
	TIPO_PUNTAJE	Char(1)	Y	Si es numérico o porcentual.
	RECALIFICAR	Char(1)	N	Puede ser recalificado o no.

Tabla:SGE_DEPENDENCIAS: Dependencias involucradas en el proceso de evaluación

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador único de la tabla
	CODIGO_DEPENDENCIA	Varchar2(20)	N	Código de la dependencia
	TIPDEP_ID	Number(22)	N	Tipo facultad o tipo escuela
	DEPEND_ID_PADRE	Number(22)	N	Identificador de la dependencia padre
	DESCRIPCION	Varchar2(100)	N	Descripción de la dependencia
	ABREVIATURA	Varchar2(15)	N	Abreviatura de la dependencia
	ESTAACTIVO	Varchar2(1)	Y	Borrado lógico del registro.
	CREADO	Varchar2(20)	Y	Usuario que crea el registro
	FECHA_CREACION	Date(7)	Y	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	JEFE_PERSONA_ID	Varchar2(15)	Y	Identificador del jefe de dependencia
	PROEVA_ID	Number(22)	Y	Identificador de proceso de evaluación

Tabla:SGE_DETALLE_ASIGNATURAS: Almacena detalle de cada asignatura

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
----	---------	-----------	-------	-------------



S	ID	Number(22)	N	Identificador único de la tabla
S	PERSON_ID	Varchar2(15)	N	Identificador del estudiante
	PERLEC_ID	Number(22)	N	Identificador del período lectivo
	CARRERA_ID	Number(22)	N	Identificador de la carrera que cursa el estudiante
	ASIGNA_ID	Number(22)	N	Identificador de la asignatura que toma el estudiante
	GRUPO_ID	Number(22)	N	Identificador que indica el grupo al cual pertenece el estudiante
	PROFESOR_ID	Varchar2(15)	N	Identificador que indica que profesor dicta cierta asignatura a un estudiante,
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	ESTAACTIVO	Char(1)	N	Borrado lógico del registro
	FACULTAD_ID	Number(22)	N	Identificador de la facultad estudiante
	ESCUELA_ID	Number(22)	Y	Identificador de la escuela
	ESROTATIVA	Varchar2(1)	Y	Indica si la asignatura es o no rotativa
	ASIGNA_MOSTRAR	Number(22)	Y	Asignatura de la malla del estudiante Puede ser diferente de asigna id

Tabla:SGE_DETALLE_DEPDOCENCIA: Detalle de docencia

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	NUMBER(22)	N	Id de la tabla
	PERSON_ID	VARCHAR2(15)	N	Id de la persona
	PERLEC_ID	NUMBER(22)	N	Id del periodo lectivo
	AREEVA_ID	NUMBER(22)	N	Id de la área de evaluación
	ESTAACTIVO	VARCHAR2(1)	Y	Estado activo s/n
	FACULTAD_ID	NUMBER(22)	N	Id de la facultad
	ESCUELA_ID	NUMBER(22)	N	Id de la escuela
	CREADO	VARCHAR2(20)	N	Usuario que crea el registro
	FECHA_CREACION	DATE(7)	N	Fecha de creación del registro
	ACTUALIZADO	VARCHAR2(20)	Y	Usuario que actualiza el registro
	FECHA_ACTUALIZACION	DATE(7)	Y	Fecha de actualización del registro
	FACULTAD_ORI	NUMBER(22)	Y	Facultad de origen
	ESCUELA_ORI	NUMBER(22)	Y	Escuela de origen

Tabla:SGE_DETALLE_DOCENCIA: Almacena información con detalle de docencia

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador único de la tabla
	PERSON_ID	Varchar2(15)	N	Identificador del docente
	PERLEC_ID	Number(22)	N	Identificador del período lectivo.
	AREEVA_ID	Number(22)	N	Identificador del área de evaluación.



NUMHORAS	Number(22)	N	Número de horas que dicta clases.
ESTAACTIVO	Varchar2(1)	Y	Borrado lógico del registro
FACULTAD_ID	Number(22)	N	Identificador que indica la facultad en la que da cátedra
ESCUELA_ID	Number(22)	N	Identificador que indica la escuela en la que da cátedra
CREADO	Varchar2(20)	N	Usuario que crea el registro
FECHA_CREACION	Date(7)	N	Fecha de creación del registro
ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro

Tabla:SGE_DETALLES_EVALUACION: Almacena información de los detalles de la evaluación

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificar de la tabla
	CUESTI_ID	Number(22)	Y	Identificador del cuestionario
	PREGUN_ID	Number(22)	Y	Identificador de la pregunta
	AREEVA_ID	Number(22)	Y	Identificador del área de evaluación.
	TIPRES_ID	Number(22)	Y	Identificador del tipo de respuesta
	OPCRES_ID	Number(22)	Y	Identificador de respuesta seleccionada
	VALOR	Varchar2(4000)	Y	Identificador que indica el campo de la pregunta si es un campo texto.
	CALIFICACION	Number(22)	Y	Identificador del valor en la pregunta
	ESTADO	Number(22)	Y	Indica si está en vigencia o está en estado recalificado
	CREADO	Varchar2(20)	Y	Usuario que crea el registro
	FECHA_CREACION	Date(7)	Y	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	ESTAACTIVO	Char(1)	Y	Borrado lógico del registro
	EVALUA_ID	Number(22)	Y	Identificador de la evaluación

Tabla:SGE_EVALUACION: Almacena información de la evaluación

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de la tabla
	PERLEC_ID	Number(22)	N	Identificador del período el lectivo.
	PROEVA_ID	Number(22)	N	Identificador del proceso de evaluación.
	CUESTI_ID	Number(22)	N	Identificador del cuestionario
	CARRERA_ID	Number(22)	N	Identificador de la carrera
	ASIGNA_ID	Number(22)	N	Identificador de la asignatura que toma el estudiante que está evaluando
	GRUPO_ID	Number(22)	N	Identificador que indica el grupo de la asignatura
	EVALUADOR	Varchar2(15)	N	Identificador del evaluador.



EVALUADO	Varchar2(15)	N	Identificador del evaluado.
ESTADO	Varchar2(1)	N	Estado de la evaluación S=normal, r=recalificada, a=autogenerada
ESTAACTIVO	Varchar2(1)	N	Borrado lógico del registro
CREADO	Varchar2(20)	N	Usuario que crea el registro
FECHA_CREACION	Date(7)	N	Fecha de creación del registro
ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
SECUENCIA	Number(22)	Y	Secuencia
FACULTAD_ID	Number(22)	N	Identificador de la facultad
ESCUELA_ID	Number(22)	N	Identificador de la escuela
TIPOEVALUADOR_ID	Number(22)	Y	Tipo de evaluador.
TIPOEVALUADO_ID	Number(22)	Y	Tipo de evaluado.

Tabla:SGE_FUNCIONES_AREAS: Almacena las funciones de la área

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
	AREAEVA_ID	Number(22)	N	Id del área de evaluación
	FUNCION_ID	Number(22)	N	Id de la tabla función
	ID	Number(22)	N	Id de la tabla
	ESTAACTIVO	Varchar2(1)	N	Estado activo s/n
	CREADO	Varchar2(15)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del usuario
	ACTUALIZADO	Varchar2(15)	Y	Usuario que actualiza el registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro

Tabla:SGE OPCIONES RESPUESTAS: Almacena las opciones de respuesta de las evaluaciones

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de la tabla
	TIPRES_ID	Number(22)	N	Tipo de respuesta.
	DESCRIPCION	Varchar2(250)	N	Nombre de la opción de respuesta
	VALOR	Varchar2(50)	N	Valor
	ABREVIATURA	Varchar2(15)	Y	Abreviatura de opción de respuesta
	PRIORIDAD	Number(22)	N	Prioridad en el conjunto de respuestas posibles.
	ORDEN	Number(22)	N	Orden de la respuesta
	ESTAACTIVO	Char(1)	N	Borrado lógico del registro
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(4000)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro

Tabla:SGE_PERSONAS: Almacena las personas que intervienen en la evaluación

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
----	---------	-----------	-------	-------------



S	ID	Varchar2(15)	N	Identificador de la tabla persona
	PERLEC_ID	Number(22)	Y	Id del periodo lectivo
	TIPO_IDENTIFICACION	Varchar2(255)	Y	Cedula - pasaporte
	APELLIDOS	Varchar2(255)	Y	Apellidos de la persona
	NOMBRE	Varchar2(255)	Y	Nombres de la persona
	EMAIL	Varchar2(255)	Y	Email de persona
	GENERO	Varchar2(20)	Y	Masculino - femenino
	ESTADO	Char(1)	Y	Estado
	CREADO	Varchar2(20)	Y	Usuario que crea la persona
	FECHA_CREACION	Date(7)	Y	Fecha de creación de la persona
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualizó el registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización de los datos
	ESTAATIVO	Char(1)	Y	Borrado lógico del registro

Tabla:SGE_PREGUNTAS: Almacena las preguntas que forman parte de la cuestionarios

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador único de la tabla
	CODIGO	Varchar2(10)	N	Código de la pregunta.
	DEFINICION	Varchar2(255)	N	Descripción de la pregunta
	TOOLTIP	Varchar2(255)	Y	Tooltip que aparece al situarse sobre la pregunta como ayuda.
	ABREVIATURA	Varchar2(15)	Y	Descripción corta para la pregunta
	ESTAATIVO	Char(1)	N	Borrado lógico del registro
	ORDEN	Number(22)	Y	Campo que sirve para ordenar preguntas.
	VALOR	Number(22)	N	Valor que tendrá la pregunta.
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro

Tabla:SGE_PROCESOS_EVALUACION: Almacena información sobre el proceso de evaluación a realizarse

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de la tabla
	NOMBRE	Varchar2(255)	N	Nombre del proceso de evaluación.
	DESCRIPCION	Varchar2(255)	N	Descripción del proceso de evaluación.
	ABREVIATURA	Varchar2(15)	Y	Abreviatura del proceso de evaluación.
	ORDEN	Number(22)	Y	Campo que sirve para ordenar varios procesos de evaluación.
	FECHA_INICIAL	Date(7)	N	Fecha de inicio del proceso.
	FECHA_FINAL	Date(7)	N	Fecha en la que finaliza el proceso.
	ENVIGENCIA	Varchar2(1)	N	Indica si el proceso está vigente o no.



ESTAACTIVO	Varchar2(1)	N	Borrado lógico del registro
CREADO	Varchar2(20)	N	Usuario que crea el registro
FECHA_CREACION	Date(7)	N	Usuario que crea el registro
ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
PERLEC_ID	Number(22)	N	Identificador del periodo lectivo
FECHA_PRORROGA	Date(7)	Y	Fecha de prórroga del proceso.
FECHA_CIERRE	Date(7)	Y	Fecha de cierre del proceso.
FECHA_RECALIFICAINI	Date(7)	Y	Fecha de inicio de recalificaciones.
FECHA_RECALIFICAFIN	Date(7)	Y	Fecha que finaliza recalificaciones.
FECHA_RESULTINI	Date(7)	Y	Fecha de inicio de presentación de resultados a los docentes
FECHA_RESULTFIN	Date(7)	Y	Fecha que finaliza el período de presentación de resultados
CABEVA_ID	Number(22)	N	Id de la cabecera evaluación

Tabla:SGE_RESPUESTAS: Almacena información de las respuestas

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	OPCRES_ID	Number(22)	N	Identificador de la opción de respuesta a la que está asociada.
S	PREGUN_ID	Number(22)	N	Identificador de la pregunta.
	ORDEN	Number(22)	Y	Criterio de ordenamiento
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	ID	Number(22)	N	Identificador de la tabla
S	TIPRES_ID	Number(22)	N	Identificador del tipo de respuesta

Tabla:SGE_TIPOS_ACTORES: Almacena los tipos de actores que intervienen en la evaluación

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de la tabla.
	NOMBRE	Varchar2(255)	N	Nombre del tipo de actor.
	DESCRIPCION	Varchar2(255)	Y	Descripción del tipo de actor.
	ABREVIATURA	Varchar2(15)	Y	Abreviatura para el tipo de actor.
	ESTAACTIVO	Varchar2(1)	Y	Borrado lógico del registro
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	ORDEN	Number(22)	Y	Criterio de ordenamiento

Tabla:SGE_TIPOS_AREAS: Almacena los tipos de áreas

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
----	---------	-----------	-------	-------------



S	ID	Number(22)	N	Identificador de la tabla.
	DESCRIPCION	Varchar2(50)	N	Descripción del tipo de área.
	ABREVIATURA	Varchar2(15)	Y	Abreviatura del tipo de área.
	ESTAACTIVO	Char(1)	N	Borrado lógico del registro
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	Y	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	ORDEN	Number(22)	Y	Criterio de ordenamiento.

Tabla:SGE_TIPOS_CUESTIONARIOS: Almacena los tipos de cuestionarios

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de la tabla.
	NOMBRE	Varchar2(50)	N	Nombre del tipo de cuestionario.
	DESCRIPCION	Varchar2(250)	N	Descripción del tipo de cuestionario
	ESTAACTIVO	Varchar2(1)	N	Borrado lógico del registro
	ABREVIATURA	Varchar2(15)	Y	Abreviatura del tipo de cuestionario.
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	ORDEN	Number(22)	Y	Criterio de ordenamiento

Tabla:SGE_TIPOS_RESPUESTAS: Almacena los tipos de respuestas

PK	COLUMNA	TIPO_DATO	NULOS	COMENTARIOS
S	ID	Number(22)	N	Identificador de la tabla
	NOMBRE	Varchar2(50)	N	Nombre del tipo de respuesta
	DESCRIPCION	Varchar2(255)	Y	Descripción del tipo de respuesta.
	TIPODATO	Varchar2(100)	N	Tipo de dato del tipo de respuesta. String-integer-date
	TAMANIO	Number(22)	N	Tamaño máximo del tipo de respuesta
	ABREVIATURA	Varchar2(15)	Y	Abreviatura del tipo de respuesta
	ESTAACTIVO	Char(1)	N	Borrado lógico del registro
	CREADO	Varchar2(20)	N	Usuario que crea el registro
	FECHA_CREACION	Date(7)	N	Fecha de creación del registro
	ACTUALIZADO	Varchar2(20)	Y	Usuario que actualiza un registro
	FECHA_ACTUALIZACION	Date(7)	Y	Fecha de actualización del registro
	NUMEROCOMPONENTES	Number(22)	N	Número de opciones de respuesta
	TIPOCOMPONENTE	Varchar2(255)	N	Indica el tipo de componente
	ORDEN	Number(22)	N	Criterio de ordenamiento.

Fuente: Elaboración propia