

UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE QUITO

CARRERA:

INGENIERÍA DE SISTEMAS

Trabajo de titulación previo a la obtención del título de:

INGENIEROS DE SISTEMAS

TEMA:

**CREACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE
NEGOCIOS PARA LA GENERACIÓN DE REPORTES CON
INFORMACIÓN RELEVANTE DEL PROCESO DE TITULACIÓN
DE LOS ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE
SISTEMAS DE LA UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO.**

AUTORES:

LUIS JHONNY BERMEO TORRES

EDISON ANDRÉS SIMBAÑA VELASTEGUI

TUTOR:

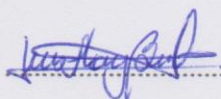
FRANKLIN EDMUNDO HURTADO LARREA

Quito, febrero del 2020

CESIÓN DE DERECHOS DE AUTOR

Nosotros Bermeo Torres Luis Jhonny y Simbaña Velastegui Edison Andrés, con documento de identificación N° 171466805-8 y 172125335-7, manifestamos nuestra voluntad y cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del trabajo de titulación intitulado: "CREACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA GENERACIÓN DE REPORTES CON INFORMACIÓN RELEVANTE DEL PROCESO DE TITULACIÓN DE LOS ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD POLITÉCNICA SALESIANA SEDE QUITO.", mismo que ha sido desarrollado para optar por el título de Ingeniero de Sistemas en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en nuestra condición de autores nos reservamos los derechos morales de la obra antes citada. En concordancia, suscribimos este documento en el momento que hacemos entrega del trabajo final en digital a la Biblioteca de la Universidad Politécnica Salesiana.



Luis Jhonny Bermeo Torres

C.I: 1714668058



Edison Andrés Simbaña Velastegui

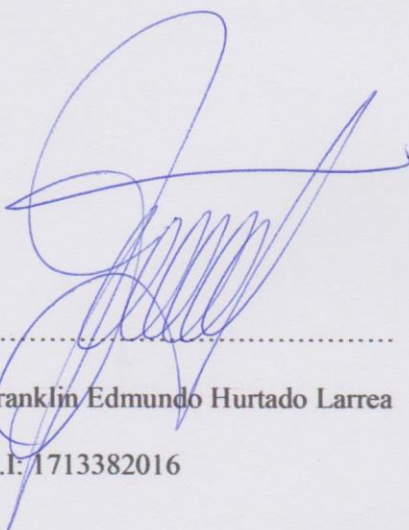
C.I: 172125335-7

Quito, febrero del 2020

DECLARATORIA DE COAUTORÍA DEL DOCENTE TUTOR/A

Yo declaro que bajo mi dirección y asesoría fue desarrollado el Proyecto Técnico, "CREACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA GENERACIÓN DE REPORTES CON INFORMACIÓN RELEVANTE DEL PROCESO DE TITULACIÓN DE LOS ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD POLITÉCNICA SALESIANA SEDE QUITO.", realizado por Bermeo Torres Luis Jhonny y Simbaña Velastegui Edison Andrés, obteniendo un producto que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana, para ser considerados como trabajo final de titulación.

Quito, febrero del 2020



.....

Franklin Edmundo Hurtado Larrea

C.I: 1713382016

DEDICATORIA

A mis padres Víctor y Olga que han sido mi motor fundamental brindándome su apoyo, cariño y confianza en cada paso que he dado para conseguir mis metas y son los principales gestores de esta instancia.

A mis hermanos con los que he compartido todo y forman parte fundamental de mi desarrollo como persona.

Finalmente, a todas las personas que han formado parte de mi vida directa o indirectamente con sus enseñanzas y experiencias.

Edison

DEDICATORIA

A la memoria de mi padre, quien me apoyó a lo largo de mi carrera y con su ejemplo me demostró que la perseverancia, la constancia y el esfuerzo siempre tienen su recompensa.

A mi madre por haberme brindado su cariño y apoyo incondicional, junto con el resto de mi familia quienes han estado presentes en todo momento y han compartido conmigo su generosidad y aprecio.

A mi esposa quien ha sido el soporte de mi vida en estos últimos años, mi compañera sentimental, y quien me ha alentado a continuar con mi carrera profesional

Jhonny

AGRADECIMIENTO

Agradecemos a la carrera de Ingeniería de Sistemas de la Universidad Politécnica Salesiana, sede Quito, por hacernos crecer a nivel profesional y a nivel humano, y por habernos permitido compartir gratos momentos que llevaremos siempre en la memoria.

Al Ing. Franklin Hurtado Larrea, por su predisposición en brindarnos su guía y asesoramiento durante el transcurso y finalización del presente proyecto.

A Dios por habernos bendecido durante el trayecto de la carrera y habernos iluminado para llevar a cabo este proyecto.

A nuestros amigos y compañeros con quienes hemos atravesado esta etapa de nuestra vida.

Edison Simbaña Velastegui

Jhonny Bermeo Torres

ÍNDICE

INTRODUCCIÓN	1
Problema de estudio	1
Justificación.....	2
Objetivos	3
Objetivo general:.....	3
Objetivos específicos:	3
Marco metodológico	4
Modelo Cascada para el desarrollo del software.....	4
Metodología de Kimball para el desarrollo del data mart.....	5
Metodología para el testeo del data warehouse.....	8
Marco referencial	10
Unidad de Titulación.....	10
Proceso de Titulación.....	10
Consejo de Carrera de Grado	11
CAPÍTULO 1	12
1.Marco teórico	12
1.1.Uso de BPMN para soluciones de BI.....	12
1.2.Metodología Kimball para datawarehousing	14
1.3.Modelamiento Multidimensional de base de datos	16
1.3.1.Modelo Copo de Nieve	16
1.3.2.Data Warehouse	16

1.3.3.Data Mart	17
1.3.4.Cubos OLAP	17
1.4.Metodología para el testeo del data warehouse.....	18
1.5.Minería de Datos	19
1.5.1.Extracción de datos relevantes de grandes volúmenes de datos	20
1.6.Proceso ETL.....	21
1.6.1.Extracción	23
1.6.2.Transformación	23
1.6.3.Carga... ..	24
1.7.Pentaho Community Edition	25
1.7.1.Schema Workbench	26
1.7.2.Saiku Analytics	26
1.8.MySQL.....	26
1.9.JSF.....	26
1.10.Tesseract OCR	27
1.11.Apache POI.....	27
CAPÍTULO 2.....	28
2.Diseño y construcción	28
2.1.Especificación de requerimientos	28
2.2.Descripción de la solución	29
2.3.Descripción de los actores.....	31
2.4.Análisis y diseño del módulo ETL.....	34

2.4.1.Extracción	34
2.4.2.Diseño del ETL	35
2.5.Análisis y diseño del data mart	37
2.5.1.Dimensiones.....	40
2.5.2.Tabla de hechos.....	41
2.5.3.Medidas	42
2.5.4.Análisis ROLAP	43
CAPÍTULO 3.....	44
3.Explotación	44
3.1.Descripción del proceso de ingreso de datos al ETL	45
3.1.1.Módulo Inscripción	45
3.1.2.Módulo Curso de Actualización.....	48
3.1.3.Ingreso de Temas de Titulación	50
3.1.4.Ingreso de lectores de planes de titulación.....	52
3.1.5.Ingreso Inicio de Titulación (Estudiantes inscritos y aprobados)	54
3.1.6.Ingreso estudiantes Curso de Actualización	61
3.2.Descripción del proceso de ingreso de datos al data mart	63
3.2.1.Proceso de ingreso de temas, alumnos y lectores asignados en la Unidad de Titulación.....	64
3.2.2.Proceso de ingreso de datos históricos.....	65
3.2.3.Proceso ETL desde las fuentes primarias de datos (Resoluciones)	71
3.4.Levantamiento de la solución en el servidor.....	79

3.5.Descubrimientos.....	82
GLOSARIO.....	92
REFERENCIAS BIBLIOGRÁFICAS.....	94
ANEXOS.....	96

ÍNDICE DE FIGURAS

Figura 1: Diagrama que describe la estructura del sistema.....	5
Figura 2: Ciclo de vida dimensional del negocio.....	15
Figura 3:Diagrama que describe las clases que componen el módulo ETL	36
Figura 4: Diagrama físico que describe la estructura de las tablas dimensionales	40
Figura 5: Diagrama físico de la tabla de hechos	41
Figura 6: Diagrama físico que muestra la estructura copo de nieve de la base de datos multidimensional.....	42
Figura 7: Diagrama que describe el flujo de información a través del sistema	44
Figura 8: Imagen que muestra el formato del archivo para el ingreso de Inscritos a UT	47
Figura 9: Imagen que muestra la interfaz para el Ingreso de alumnos inscritos en la UT	48
Figura 10: Imagen que muestra el formato del archivo para el ingreso estudiantes del curso de actualización a la UT	49
Figura 11: Imagen que muestra la interfaz de ingreso de estudiantes del curso de actualización a la UT.....	50
Figura 12: Formato de ingreso de temas de proyectos.....	51
Figura 13: Imagen que muestra el formato del archivo para la asignación de tema y lector.....	53
Figura 14: Imagen que muestra la interfaz de asignación del tema y lector	54
Figura 15: Imagen que muestra la interfaz de login del estudiante.....	55
Figura 16: Imagen que muestra la interfaz de ingreso de documentos de Inicio del proceso de titulación	56
Figura 17: Interfaz Gestión del proceso de titulación	58

Figura 18: Imagen que muestra la interfaz de gestión de cambios de tutor, lector o tipo de proyecto.....	60
Figura 19: Imagen que muestra la Interfaz de finalización del proceso de titulación	61
Figura 20: Imagen que muestra la interfaz para el login de los estudiantes de actualización.....	62
Figura 21: Imagen que muestra la base de datos primaria de la UT	63
Figura 22: Estructura de la base de datos primaria de la UT	66
Figura 23: Imagen que muestra la base de datos primaria de la UT una vez que fue procesada (estandarizada) manualmente.....	67
Figura 24: Imagen que muestra los comandos SQL necesarios para el ingreso de datos en la BDD almacenada en el gestor MySQL.....	68
Figura 25: Imagen que muestra la creación de la conexión Pentaho-MySQL con la herramienta Data Source Wizard	69
Figura 26: Imagen que muestra la interfaz de la herramienta Schema Workbench...	70
Figura 27: Imagen que muestra la creación del reporte Historial de alumnos, con la herramienta Saiku Analytics	71
Figura 28: Imagen que muestra el reporte fijo del historial de los alumnos.....	75
Figura 29: Imagen que muestra el reporte del porcentaje de alumnos por opción de titulación y por periodo	76
Figura 30: Imagen que muestra el reporte del número de temas propuestos por periodo y por tipo de proyecto de titulación	77
Figura 31: Imagen que muestra la página de inicio de la herramienta Saiku Analytics	78
Figura 32: Panel de creación de reportes de la herramienta Saiku Analytics	78
Figura 33: Construcción del proyecto	79

Figura 34: Directorio de localización de archivo .war.....	80
Figura 35: Gestor de aplicaciones web del servidor Tomcat	80
Figura 36: Configuración archivo server.xml de tomcat	81
Figura 37: Reporte de descubrimiento: "Porcentaje de alumnos que aprobaron el examen complejo por periodo"	82
Figura 38: Reporte de descubrimiento: "Cantidad de veces que un profesor fue elegido lector en varios periodos"	83
Figura 39: Reporte de descubrimiento: "Estudiantes que han ido a Curso de Actualización en determinado periodo."	84
Figura 40: Reporte de descubrimiento: "Estudiantes que han solicitado 2da prorroga en determinado periodo"	85
Figura 41: Diagrama físico anterior del data mart	86
Figura 42: Modelo físico optimizado del data mart	87
Figura 43: Reporte de descubrimiento: " Estudiantes que han cambiado de tutor en su proceso de titulación"	88

ÍNDICE DE TABLAS

Tabla 1: Descripción del servidor.	7
Tabla 2: Actor Unidad de Titulación.	31
Tabla 3: Actor Tesorería de sede.	32
Tabla 4: Actor Secretaría de sede.....	32
Tabla 5: Actor Consejo de Carrera.....	33
Tabla 6: Actor Dirección de Carrera.	33
Tabla 7: Actor Estudiante.....	34
Tabla 8: Matriz Proceso/Dimensión para el Proceso de Titulación.	38
Tabla 9: Involucrados y tablas.	39

RESUMEN

El presente documento presenta el desarrollo de una solución de Inteligencia de Negocios para la gestión del proceso de Titulación de la carrera de Ingeniería de Sistemas sede Quito, se conoce que la información de este proceso se maneja en documentos dispersos en distintos repositorios, además de contener archivos planos de diferentes extensiones, por tal motivo el tratamiento de la información se ve aletargado para llegar a tomar decisiones. Razón por la cual es necesario un sistema de gestión de datos con la utilización de un ETL (Extract-Transform-Load), que fue desarrollado siguiendo una metodología lineal-secuencial al contar con los requisitos planteados de manera específica, juntamente con una base de datos multidimensional (data mart) que se acoge a la metodología de Kimball la cual está basada en el ciclo dimensional del negocio.

Los módulos descritos interactúan con la herramienta Pentaho para realizar consultas de la base de datos multidimensional utilizando diferentes parámetros alineados al negocio con la finalidad de brindar reportes de valor para la gestión del proceso.

ABSTRACT

This document presents the development of a Business Intelligence solution for the management of the Titling process of the Systems Engineering degree in Quito, it is known that the information of this process is handled in documents dispersed in different repositories in addition to containing files of different extensions, for this reason the treatment of information is delayed to reach decisions. Reason why a data management system is necessary with the use of an ETL (Extract-Transform-Load), which was developed following a linear-sequential methodology by having the requirements set out specifically, together with a database of multidimensional data (data mart) that is based on the Kimball methodology which is based on the dimensional business cycle.

The modules described interact with the Pentaho tool to make queries of the multidimensional database using different parameters aligned to the business in order to provide value reports for process management.

INTRODUCCIÓN

La Unidad de Titulación es el ente encargado de manejar el proceso de titulación de los estudiantes de la carrera de Ingeniería de Sistemas de la sede Quito, actualmente los documentos que se generan de este son almacenados en repositorios digitales externos y en archivos de texto planos siendo poco eficientes para su tratamiento.

Es necesario una respuesta rápida y ordenada para poder gestionar el proceso con un enfoque de mejora, esto se reflejará a través de reportes estructurados y no estructurados que estarán a disposición de la Unidad de Titulación.

En el capítulo uno se describe el sustento teórico utilizado como referencia para este proyecto el cual brinda las pautas a seguir para su correcta ejecución.

El capítulo dos se basa en el desarrollo y construcción tanto del módulo ETL como del data mart, adicional a estos se propone la utilización de un módulo de reportes que utiliza la tecnología de Pentaho.

En el capítulo tres se explica la explotación de los datos, la manera que fueron cargados en el sistema tanto datos históricos como los datos que ingresarán de manera continua.

Finalmente se obtienen reportes reflejados en la herramienta descrita que permiten conocer y descubrir nuevos indicadores que serán de utilidad para realizar una mejor gestión del proceso.

Problema del estudio

La Unidad de Titulación de la Carrera de Ingeniería de Sistemas maneja información presente en documentos (dispersos en distintos repositorios) como son: resoluciones, informes de docentes y documentos de facturación. Dicha información resulta crítica y de valor para el coordinador de la Unidad de Titulación, la Dirección de Carrera y

los estudiantes inscritos en la Unidad de Titulación, por ende, debe ser gestionada de manera correcta para la validación de las actividades que conciernen al proceso.

Actualmente la Unidad de Titulación utiliza la herramienta Excel de Microsoft, la cual no resulta lo suficientemente eficiente para el tratamiento de la información. Con esta herramienta, de manera manual, se realiza las consultas sobre el estado del proceso.

Justificación

La creación de una solución de inteligencia de negocios para gestionar la información de la Unidad de Titulación de la Carrera de Ingeniería de Sistemas de la UPS Sede Quito tiene como objetivo automatizar y agilizar la obtención, presentación y visualización.

La utilización de un data mart como un módulo de almacenamiento multidimensional brinda el enfoque previsto para el segmento de negocio que se gestionará, ya que la información pertenece a un proceso ya identificado, no es necesaria una estructura de mayor volumen de datos como un Data warehouse. La base de datos multidimensional permite la utilización de aplicaciones de procesamiento analítico en línea (OLAP) y se enfoca en responder hacia consultas más abiertas de los datos, al contrario de una base de datos relacional donde se encuentran limitaciones de utilización de consultas definidas.

El sistema definirá parámetros de entrada de la información, con esto el sistema podrá analizar y gestionar información de periodos anteriores siempre y cuando se cumpla con los formatos de ingreso de datos, optimizando las consultas acerca del proceso además de la disponibilidad inmediata de la información que apoyará a la toma de decisiones.

Objetivos

Objetivo general:

Implementar una solución de Inteligencia de Negocios para la generación de reportes basados en la información relevante del proceso de Titulación que maneja la Unidad de Titulación de la Carrera de Ingeniería de Sistemas sede Quito.

Objetivos específicos:

Analizar e identificar las fuentes de datos para una solución de Inteligencia de Negocios que apoye a la gestión y presentación eficiente de información sobre el proceso de Titulación.

Desarrollar un módulo ETL para la extracción de los datos ubicados en los distintos documentos y repositorios.

Analizar e implementar el modelo de copo de nieve de la base de datos multidimensional (data mart) para alojar los datos extraídos por el módulo ETL.

Implementar Pentaho como herramienta para la administración y visualización de reportes sobre el proceso de Titulación.

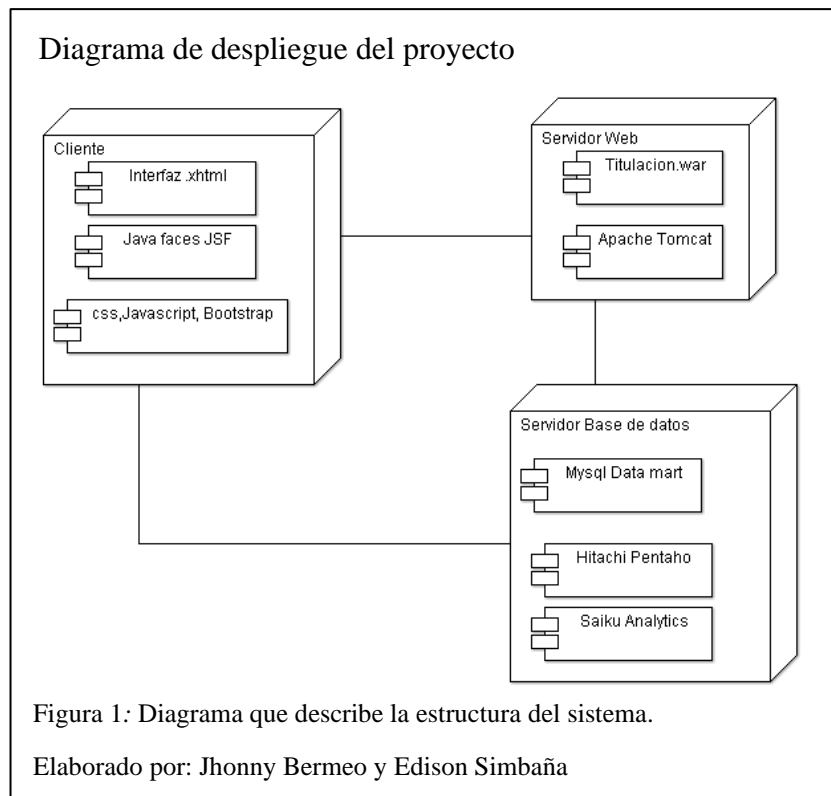
Marco metodológico

Modelo Cascada para el desarrollo del software

Para la ejecución de este proyecto se utilizó el modelo de desarrollo en cascada, puesto a que los requisitos fueron definidos de manera clara y, a pesar de los cambios realizados a lo largo del proyecto, no ameritó una metodología basada en iteraciones.

El modelo de cascada fue definido como una secuencia de actividades en el cual se puede completar una actividad antes de poder continuar con la siguiente actividad. (Pradel & Raya, 2015), se plantearon las siguientes etapas que fueron realizadas de manera secuencial:

- **Requisitos:** Se realizó una indagación sobre el Proceso de Titulación con el fin de obtener información necesaria sobre las fases del proceso, los involucrados y los parámetros que definen al proceso. A continuación, se obtuvieron los requerimientos para el desarrollo de la solución de inteligencia de negocios.
- **Análisis y diseño:** Se analizó el Proceso de Titulación establecido y en base a este análisis se diseñaron los diferentes componentes de la solución de inteligencia de negocios representados en la Figura 1. Se diseñó el data mart de la solución y el módulo ETL que la alimenta.



Módulo del usuario: Tanto para estudiantes como para la Coordinación de la Unidad de Titulación consta de interfaces definidas en formato .xhtml que implementa tecnología JSF junto con hojas de estilo, javascript y Bootstrap.

Servidor Web: que aloja el aplicativo Titulación basado en un servidor Tomcat.

Servidor de base de datos: Aloja al data mart elaborado en MySQL que se comunica con el componente de reportes y análisis Pentaho junto a Saiku analytics.

Metodología de Kimball para el desarrollo del data mart

Para el desarrollo e implementación del data mart se siguió la metodología de Kimball, la cual está basada en el ciclo dimensional del negocio (Business Dimensional Lifecycle).

La Metodología define seis fases para la construcción de la solución de inteligencia de negocio, estas fases son:

1. Planificación del proyecto: En esta fase se identifica el propósito del proyecto. Se definió que el propósito del proyecto es automatizar y agilizar la obtención, presentación y visualización de información relevante para el Proceso de Titulación de la Unidad de Titulación de la carrera de Ingeniería de Sistemas sede Quito.

Se planifico el proyecto en base al macroproceso de titulación generado en base a los diagramas BPMN (Business Process Model and Notation) del proceso de titulación (Ver anexo 1)

2. Definición de los requerimientos del negocio: Durante esta fase se identifican los factores claves que guían el negocio. Se establecieron los parámetros y factores en base al Proceso de Titulación modelado en BPMN en el Anexo 1, del cual se generó la matriz proceso/dimensión. (Ver Tabla 6)

3. Modelado dimensional: Durante esta fase se debe determinar el nivel de granularidad de la información, las dimensiones, las medidas y las tablas de hechos.

Se procedió a establecer la granularidad de las dimensiones propuestas, generando la matriz de granularidad. (Ver Anexo 2)

4. Diseño Físico: se determina la mejor estructura que se adapta al modelo dimensional previamente determinado, Kimball propone mantener los nombres establecidos en el diseño lógico y se procedió a la creación del modelo físico de la base de datos. Luego se procedió con el montaje de esta en el gestor

de base de datos MySQL. Kimball establece como entregable de este punto el diccionario de datos del modelo. (Ver diccionario de Datos Anexo 3)

5. Diseño e implementación del subsistema de Extracción, transformación y Carga: En esta fase se diseña e implementa el módulo ETL de la solución que alimenta el data mart. Se desarrolló el módulo ETL en JSF mediante la aplicación de varias librerías que permiten la extracción de datos desde diversos tipos de documentos, PDF y Excel, además de la extracción de datos a través de OCR para aquellos documentos escaneados y subidos al sistema.
6. Diseño de la arquitectura técnica: establece varios pasos para levantar la arquitectura que soportará a la solución BI, es decir, se determina el hardware y software necesario para soportar la robustez de la solución. Para el software de la solución se estableció a la herramienta Pentaho CE para la visualización y creación de reportes.

Para alojar el módulo ETL se estableció un servidor Tomcat que alojará la totalidad del módulo.

Toda la solución, desde el servidor Apache-Tomcat hasta el servidor BI Pentaho CE, serán alojados el servidor descrito en la Tabla 1:

Tabla 1: Descripción del servidor.

DESCRIPCION	NOMBRE MAQUINA VIRTUAL	RAM	DISCO	CPU
WINDOWS SERVER 2012	VM_FHURTADO_UT_BI_WS12	16 GB	400 GB	8

Nota: Descripción del servidor dispuesto por el Data Center ICC-Campus Sur.

Metodología para el testeo del data warehouse

La metodología para las pruebas, o testeo, a un data warehouse, utilizada en esta solución, es la determinada por Doug Vucevic y Wayne Yaddow, quienes proponen una metodología enfocada en los datos y basada en tres pilares fundamentales que son: el contenido de los datos extraídos por la solución, la estructura física y lógica del repositorio y la calidad de los datos almacenados en la solución, además determinan otro tipos de pruebas que ayudan a conocer el rendimiento de la solución de BI versus una solución basada en consultas relacionales.

a. Contenido de los datos extraídos por la solución

El contenido de datos provenientes del módulo ETL deben ser testeados con el fin de conocer si estos aportan la información necesaria y relevante para la solución de BI. Estas pruebas fueron realizadas con el fin de determinar la existencia de datos nulos dentro la base de datos. (Ver Pruebas de validación de datos Anexo 4)

b. Estructura

En esta fase se completó un checklist que demuestra que efectivamente el data mart está alineado al proceso de negocio seleccionado. (Ver Checklist de validación de Estructura Anexo 4)

c. Calidad

Se debe verificar y validar que la calidad de los datos obtenidos, luego del proceso ETL, estén estandarizados y sean legibles, para esta prueba se realizaron consultas de diversa índole en la BDD. (Ver pruebas de funcionalidad Anexo 4).

d. Solución BI vs Solución de consultas SQL.

Se realizaron consultas SQL sobre la base de datos con el fin obtener la misma información que la herramienta Saiku y se procedió a cronometrar las dos opciones generando una matriz de diferenciación de tiempos. (Ver Pruebas de rendimiento Anexo 4).

- **Implementación:** Durante esta fase se implementó tanto el data mart de la solución y se procedió con la implementación del módulo ETL desarrollado como una aplicación web. Conjuntamente estos dos módulos fueron implementados en el espacio designado por el Data Center ICC - Campus Sur. Esta fase está reflejada en el apartado de explotación donde se profundiza la configuración y la puesta en marcha del software.
- **Pruebas:** Una vez que la solución completa fue montada en el servidor, se procedió con las pruebas respectivas de funcionalidad y sus respectivas correcciones y validaciones.

Marco referencial

Unidad de Titulación

La Unidad de Titulación de la UPS es el ente encargado de organizar, planificar, controlar el proceso de titulación de los estudiantes que cumplan con los requisitos para obtener el título de tercer nivel. Está conformado por:

- Docente Coordinador de la Unidad de Titulación.
- Docentes de apoyo.

Se debe recalcar que los pasos del proceso dependerán del tipo de proyecto de titulación en el cual el alumno se haya inscrito. (Consejo superior de la Universidad Politécnica Salesiana, 2019).

Proceso de Titulación

El proceso de titulación dispuesto para la Unidad de Titulación de la carrera de ingeniería de sistemas sede Quito está regulada por el Reglamento del Régimen Académico del Consejo de Educación Superior CES, el cual determina las políticas con respecto al proceso de titulación, sin embargo, el Consejo Superior de la Universidad Politécnica Salesiana otorga a cada sede (Cuenca, Quito, Guayaquil) el deber de establecer los mecanismos a llevarse a cabo para el correcto proceso de titulación. (Vicerrectorado UPS, 2014)

El proceso de titulación, definido por la dirección de titulación de la carrera de ingeniería de sistemas, establece, además, los documentos que son relevantes para el proceso y aportan los datos importantes los cuales son:

- Resoluciones.
- Solicitudes.

- Réconds académicos.
- Cartas de docentes. (tutor y lector)
- Facturas.
- Cartas de auspicio. (Unidad de Titulación, 2019)

Consejo de Carrera de Grado

Se conoce como Consejo de Carrera de Grado a la entidad que gestiona a las carreras ofertadas por la Universidad Politécnica Salesiana, así como determinar los diferentes tipos de trabajo de titulación en base a lo establecido por la Unidad de Titulación según el Reglamento de Régimen Académico. (Consejo Superior UPS, 2018)

CAPÍTULO 1

1. Marco teórico

1.1. Uso de BPMN para soluciones de BI

Business Intelligence (BI)

La inteligencia de negocios (Business Intelligence), BI por sus siglas en inglés, puesto que engloba una gran cantidad de elementos, es definida desde varias perspectivas, las más relevantes son:

BI como solución tecnológica: De acuerdo con esta perspectiva, el BI es un sistema compuesto por varios componentes tecnológicos que permiten:

1. Recopilación y almacenamiento de datos.
2. Transformación de datos en información.
3. Transformación de información en conocimiento.
4. Toma de decisiones en base al conocimiento. (Shollo, 2013)

BI como un conjunto de tecnologías y procesos: Según esta perspectiva, establecida por Barbara Wixom y Hugh Watson, se puede definir al BI como un conjunto amplio de tecnologías, aplicaciones y procesos que permiten la recolección, almacenamiento, acceso y análisis de datos que ayuda en la toma de decisiones a nivel organizacional. (Wixom & Watson, 2010)

BI como proceso: Para Ralph Kimball, en su metodología de creación y desarrollo de data warehouse, BI es un proceso que consiste en obtener, recopilar, analizar y presentar información relevante para la toma de decisiones del negocio mediante la implementación, mejoramiento y mantenimiento de datawarehouse y aplicaciones BI

que consiste en módulos de extracción, transformación y carga de datos. (Kimball & Ross, 2013)

Business Process Management Notation (BPMN)

La Notación de Modelado de Procesos de negocio (Business Process Management Notation) BPMN, por sus siglas en inglés, es una forma estandarizada de representar procesos de negocio para propósitos descriptivos de alto nivel, así como para entornos de software orientados a procesos. Además, involucra a las actividades ordenadas de manera secuencial para representar los objetivos centrales de una organización. (White & Miers, 2016)

La notación de administración del proceso de negocios (BPMN) provee el modelado de los procesos que se llevan a cabo dentro de un negocio, con el objetivo de comunicar y ayudar a comprender mejor la referencia global del negocio. (White & Miers, 2016)

Definición del proceso de negocio a través de BPMN

Según Stephen White, la meta de la notación BPM es permitir la capacitación de los usuarios finales, brindándoles una notación simple y acordada que resume y comunica el proceso de negocio de una organización, siempre y cuando este proceso estuviera bien definido, por ende, un usuario final es capaz de entender el proceso, los actores y el flujo del proceso de negocio partiendo desde el modelo BPMN. (White & Miers, 2016)

Uso de BPMN para el levantamiento de soluciones BI

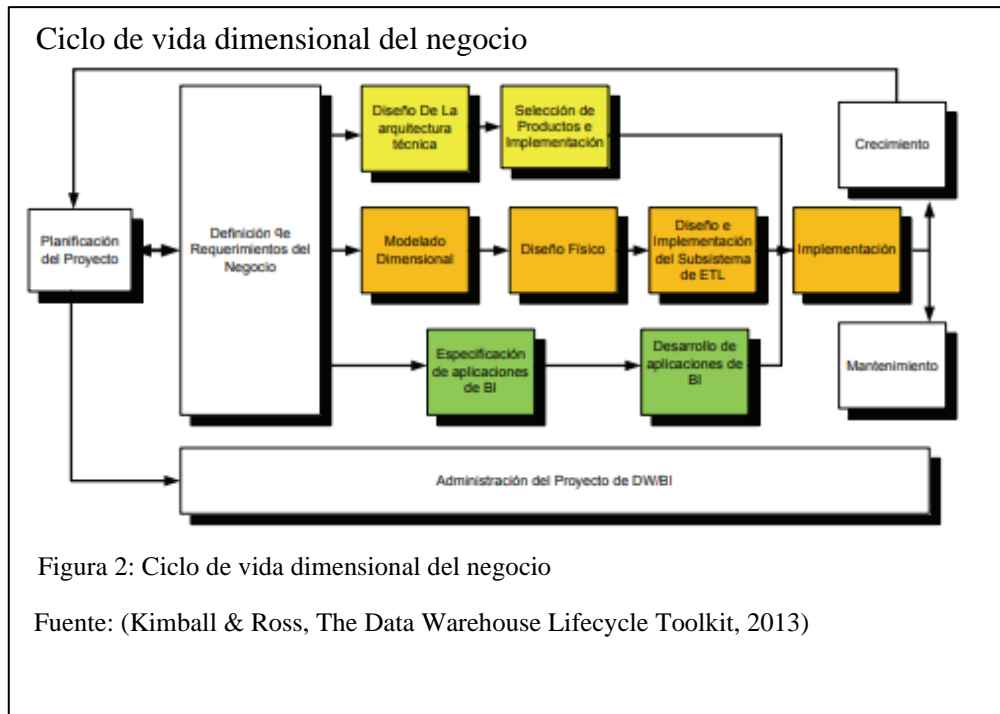
Como lo explica Tobías Bucher, el uso de BPMN para soluciones de BI no es un tema nuevo puesto que la documentación del flujo de actividades que se realiza en una organización favorece a una sencilla integración con una solución de inteligencia de negocios. Esta integración entre el proceso de la organización y la solución BI surge

debido a que varios procesos operativos generan datos que se almacenaran en grandes Almacenes de Datos y directamente podrían ser analizados por sistemas de Inteligencia de Negocios. Si los procesos operativos están debidamente documentados, específicamente en notación BPMN, es mucho más sencillo levantar una solución de BI puesto que se tiene documentado la fase del proceso en la cual se generan determinados tipos de datos y se puede analizar el impacto de estos a nivel organizacional para su posterior inclusión en la solución BI. (Bucher & Gericke, 2009)

1.2. Metodología Kimball para datawarehousing

La metodología de Kimball para construir un data warehouse está basada en el ciclo dimensional del negocio (Business Dimensional Lifecycle), como se muestra en la Figura 2, el cual se fundamenta en tres principios:

1. Añadir valor al negocio.
2. Estructurar las dimensiones en base a los involucrados en el proceso de negocio.
3. Desarrollar la solución de manera iterativa presentado valor con cada entregable.



Además, la metodología define seis fases para la construcción de la solución de inteligencia de negocio, estas fases son:

1. Planificación del proyecto: identifica el alcance de la solución y el propósito de esta.
2. Definición de los requerimientos del negocio: se identifican los factores claves que guían el negocio
3. Modelado dimensional: se determina las dimensiones para cada involucrado y se especifica la granularidad de los datos
4. Diseño Físico: se determina la mejor estructura que se adapta al modelo dimensional previamente determinado.
5. Desarrollo de la presentación de datos: se desarrolla el módulo de extracción, transformación y carga (Extract Transform & Load), ETL por sus siglas en ingles.
6. Diseño de la arquitectura técnica: establece varios pasos para levantar la arquitectura que soportará a la solución BI, es decir, se determina el

hardware y software necesario para soportar la robustez de la solución.
(Kimball & Ross, 2013)

1.3. Modelamiento Multidimensional de base de datos

El modelo multidimensional de base de datos (Multidimensional Database), MDB por sus siglas en inglés, es un modelo que permite la identificación de los procesos de negocio y sus involucrados, de manera que permite observar, de manera clara, como interactúan entre ellos, basados en el proceso de negocio.

Se diferencia del modelo entidad relación en la manera de organizar los datos provenientes de las fuentes, esto permite una flexibilidad de diseño que logra un mejor desempeño al momento de visualizar consultas. (Kimball & Ross, 2013)

1.3.1. Modelo Copo de Nieve

El modelo copo de nieve, o snowflake, es un modelo que se aplica tanto a los data warehouse como a los data mart y permite que las tablas dimensionales guarden relación entre sí, de manera que se pueda especificar las relaciones jerárquicas existentes dentro del data mart, evitando así la redundancia dentro de la base de datos gracias a la normalización de las tablas. Presenta una tabla de hechos central la cual se alimenta de los datos de las dimensiones a su alrededor, al igual que la mayoría de los modelos de data mart.

1.3.2. Data Warehouse

Según la definición de Kimball, un data warehouse es una base de datos corporativa, desnormalizada no modelada en base a los sistemas transaccionales, que es alimentada por todos los departamentos que conforman la organización y usualmente están modelados en base a esquemas multidimensionales como son los modelos de estrella o de copo de nieve.

1.3.3. Data Mart

Kimball define los data mart como una versión más pequeña de los data warehouse. Los denomina como data warehouse departamentales, y hace énfasis en que poseen un nivel de granularidad mayor puesto que solo se enfocan en un área, o proceso, de la organización. Por lo cual el mismo Kimball establece que la única diferencia existente entre un data mart y un data warehouse radica en su alcance (uno tiene el alcance empresarial y el otro tiene un alcance departamental)

1.3.4. Cubos OLAP

Los cubos de proceso analítico en línea (On-line Analytical Processing), OLAP por sus siglas en inglés, son el resultado de la organización del modelo multidimensional de una base de datos y está definido en base a las necesidades del negocio y permiten el procesamiento de grandes cantidades de datos. Los cubos OLAP están clasificados en base al tipo de Base de Datos sobre el cual están implementados:

1.3.4.1. Cubos MOLAP (Multidimensional Online Analytical Processing)

Los cubos MOLAP son cubos OLAP cuyo análisis se realiza sobre una estructura multidimensional en la cual reposan la información y sus agregaciones. Las consultas sobre esta estructura no se realizan directamente en la base de datos, sino desde una réplica de esta, por lo cual se requiere una tarea de replicación de los datos de la fuente para generar un modelo físico aparte, con el objetivo que este nuevo modelo no interfiera con la operabilidad.

1.3.4.2. Cubos ROLAP (Relational Online Analytical Processing)

Se conoce como cubos ROLAP a aquellos cubos OLAP cuyo análisis se realiza sobre una estructura de base de datos relacional de tipo lógico, es decir, directamente sobre la base de datos fuente de este modo logra evitar la redundancia de datos y evita la creación de tareas adicionales sobre la base para la replicación de datos en un modelo físico. (Kimball & Ross, 2013)

1.4. Metodología para el testeo del data warehouse

La metodología para las pruebas, o testeo, a un data warehouse, utilizada en esta solución, es la determinada por Doug Vucevic y Wayne Yaddow, quienes proponen una metodología enfocada en los datos y basada en tres pilares fundamentales que son: el contenido de los datos extraídos por la solución, la estructura física y lógica del repositorio y la calidad de los datos almacenados en la solución, además determinan otro tipos de pruebas que ayudan a conocer el rendimiento de la solución de BI versus una solución basada en consultas relacionales.

a. Contenido de los datos extraídos por la solución

El contenido de datos provenientes del módulo ETL deben ser testeados con el fin de conocer si estos aportan la información necesaria y relevante para la solución de BI y verificar si estos datos no presentan inconsistencias o datos nulos durante la extracción.

b. Estructura

La estructura del repositorio o BDD está alineada al proceso de negocios para el cual se está desarrollando la solución. Los autores proponen realizar un checklist y consultar con los involucrados sobre si el data mart está alineado al proceso de negocio.

c. Calidad

Se debe verificar y validar que la calidad de los datos obtenidos, luego del proceso ETL, estén estandarizados y sean legibles para el usuario final. Esto se logra con pruebas de validación de datos a nivel de los reportes, Vucevic hace énfasis en esta fase pues se debe validar, junto con los interesados, si estos datos obtenidos en los reportes son legibles y entendibles.

d. Solución BI vs Solución de consultas SQL

Se debe verificar que la solución BI tenga un mejor rendimiento que una solución SQL basada en consultas, para lo cual se toma en cuenta el tiempo que ambas se tardan en realizar una misma consulta. Se debe además tomar en cuenta que las dos arrojen un mismo resultado o, como mínimo, similar debido a que los gestores de BDD tienen limitaciones en la capa de presentación de los datos, algo que las soluciones de BI no. (Vucevic & Yaddow, 2012)

1.5. Minería de Datos

Se define como Minería de datos (data mining) al conjunto de técnicas y tecnologías que permiten la exploración de repositorios de datos de gran tamaño con el fin de obtener información al de manera automática o semiautomática. (Han, Kamber, & Jian, 2012)

La minería de datos no solo se limita a obtener datos de repositorios convencionales (base de datos relacionales o transaccionales) sino también a archivos, los cuales pueden ser de texto o inclusive archivos multimedia y pueden provenir de repositorios más grandes aún como son repositorios transaccionales de negocio, repositorios de datos científicos o médicos, repositorios de datos sensoriales, etc. (Zaiane, 1999)

1.5.1. Extracción de datos relevantes de grandes volúmenes de datos

Basado en el proceso de minería de datos que propone Lorenzo Martínez, se define a la minería de datos como un conjunto de pasos para la obtención de datos relevantes desde bases de datos con grandes cantidades de información y su posterior análisis mediante la aplicación de herramientas de Proceso de Análisis en Línea.

Las técnicas de minería de datos son variadas y dependen del tipo de información que se desea obtener y del tipo de repositorio del cual se está minando datos. Las más utilizadas, como Lorenzo Martínez señala en su artículo, son:

1.5.1.1. Umbral

Se conoce como umbral o umbrales, a la técnica que consiste en colocar un rango, definiendo el máximo y mínimo valor tolerado, para que el algoritmo informe cuando el valor de un dato esta por fuera de estos límites. Se debe tomar en cuenta que los valores deben corresponder a un registro periódico para realizar el correcto análisis de este.

1.5.1.2. Tendencia

Este algoritmo permite obtener datos y observar sus variaciones considerables, en un periodo de tiempo, y un dato predecesor con el fin de poder realizar proyecciones en el tiempo.

1.5.1.3. Patrones frecuentes

Mediante este algoritmo se puede establecer una regla general al encontrar un patrón que se repita varias veces dentro de un rango de tiempo determinado, de manera que se registra como frecuente.

1.5.1.4. Cluster

Esta técnica permite la clasificación en categorías, o clusteres, a un determinado grupo de datos mediante algoritmos de agrupación, parte de la premisa de los patrones salvo que, a diferencia de las anteriores técnicas, no requiere un periodo de tiempo como mandatorio para realizar la minería. (Martinez, 2011)

1.6. Proceso ETL

El proceso de extracción, transformación y carga (Extract, Transform & Load), ETL por sus siglas en inglés, consiste en una serie de procedimientos para obtener los datos de un repositorio, limpiarlos o estandarizarlos (transformarlos) y cargarlos en la base de datos de negocio. Kimball considera que el desarrollo de este módulo consume la mayor parte del tiempo y esfuerzo dentro de la construcción del entorno de la solución de BI, esto debido varios factores entre los que destacan los requisitos de la solución, la realidad de los datos de origen y las ventanas de procesamiento.

Requisitos de la solución: Se entiende por los requisitos de la solución al alcance que tendrá la solución de BI y es un factor clave en la construcción tanto del módulo ETL como en la solución en general, puesto que se debe entender únicamente el proceso de negocio y sus tareas relevantes para la construcción de la solución. Además, se debe tener presente el contenido de información que los usuarios necesitan conocer para la toma de decisiones.

Realidad de los datos de origen: Con este factor Kimball se refiere al tipo de repositorio que provee los datos que se desea extraer, puesto que este tipo de repositorios nunca, o casi nunca, se encuentran estandarizados, además de que los datos almacenados en él rara vez cumplen con un patrón lógico y ordenado con el cual se pueda trabajar. Señala, además, que muchas organizaciones tienen fuentes de datos distribuidas, es decir se trabaja con más de una fuente de datos, lo que conlleva a

determinar si en todas estas fuentes se maneja la información de una misma manera y si se trabaja con estándares para tratar las diversas fuentes de datos como si fuera una sola.

Ventanas de procesamiento: Kimball señala este factor debido a que el módulo ETL puede interferir con el normal desempeño del sistema informático de la organización, al consumir recursos de procesamiento y memoria para ejecutar sus tareas, además hace hincapié en la latencia de datos, que es la velocidad con la que el sistema debe entregar los datos del sistema fuente a los usuarios finales, esto podría verse afectado por el alto procesamiento del módulo.

Una de las consideraciones que Kimball también propone, pero no lo señala como un factor decisivo, es el enfoque que se adopta al crear el módulo ETL, ya que, si no se tiene claro algún proceso, o se desconoce el origen y destino de los datos, puede darse resultados variados como un desorden de las tablas, desorden de procesos, problemas que pueden surgir al no seguir un enfoque estructurado para el desarrollo del ETL. (Kimball & Ross, 2013)

El proceso ETL se puede desglosar no solo en sus tres subprocesos principales (Extracción, transformación y carga) sino también en tareas, Kimball hace énfasis en esta parte puesto que el módulo ETL es el encargado de alimentar la solución de BI, por ende, desglosa a cada subproceso de la siguiente manera:

1.6.1. Extracción

El subproceso inicial del módulo el cual se encarga de obtener los datos de la fuente y transferirlos a un entorno donde el resto del ETL pueda operar con ellos, inicialmente este subproceso va a extraer datos primitivos y está conformado por las siguientes tareas:

- a) Aislamiento de datos: Consiste en diferenciar el dato de importancia del resto de los datos en la fuente, esto se logra mediante el patrón que debe seguir el módulo para la obtención del dato.
- b) Perfilamiento de datos: Realizar un perfilamiento de datos es verificar la calidad y consistencia de los datos en la fuente, es decir, que de encontrarse anomalías en los datos estas puedan ser tratadas y de esta manera puedan ser extraídas sin mucha complejidad.
- c) Validación el destino de los datos: Validar el destino de los datos incluye la revisión de los datos provenientes de la fuente y asegurarse de que sean del mismo tipo que el destino.
- d) Validación de operación de datos: Las operaciones entre datos deben ser tomadas en cuenta de manera que se pueda diferenciar si al extraer un dato este debe ser insertado en el destino (INSERT INTO...) o es un dato que indica una actualización de algún otro dato en el destino (UPDATE).

1.6.2. Transformación

La transformación de datos surge inmediatamente tras la extracción de estos, puesto que el anterior subproceso ayuda con varios problemas de formato y estándar de los datos, la transformación debe continuar el proceso mediante las siguientes tareas:

- a) **Parseo (Parsing):** Mediante esta tarea se realiza un análisis “lingüístico” sobre los datos extraídos en bruto, con la finalidad de identificar caracteres, o cadenas de caracteres, que puedan presentar un problema a la hora de ingresar el dato en la base.
- b) **Corrección sintáctica:** Una vez que se han identificado los caracteres especiales dentro de los datos, se procede con el reemplazo de estos, de manera que el dato final pueda ser entendido por el sistema y el usuario final, y no involucre una pérdida de coherencia con el dato de origen. Kimball señala como ejemplo la corrección en los caracteres de índole latina como la ‘ñ’.
- c) **Estandarización:** Finalmente, y con todos los caracteres especiales reemplazados, esta tarea lleva a los datos a una estandarización general, lo más sencillo puede ser que los caracteres alfabéticos se encuentren en su totalidad como mayúsculas o minúsculas y sean establecidos como una cadena de caracteres, al igual que las cadenas alfanuméricas, en cuanto a los caracteres numéricos deben ser establecidos como tales (Enteros o de punto flotante, etc...)

1.6.3. Carga

La carga es el subproceso más crítico puesto que si alguno de los subprocesos anteriores no se llevó a cabo de manera correcta este último subproceso no podrá cargar la información y se producirá un error durante su ejecución. Las tareas que componen este subproceso son:

- a) Comprobación: Durante esta tarea se comprueba que los datos obtenidos de la fuente coincidan con las tablas y columnas que componen el data warehouse, y no se esté enviando datos de una dimensión a otra muy distinta.
- b) Carga de dimensiones: Esta tarea carga los datos limpios en las distintas dimensiones que componen el data warehouse, se basa en la definición de la Validación de datos que ocurre durante el proceso de Extracción.
- c) Carga de la tabla de hechos: La tarea final que determina el éxito del proceso de carga, durante esta tarea la tabla de hechos es cargada únicamente con las claves primarias de las dimensiones y pueden diferenciarse dos tipos de carga para la tabla de hechos
 - i. Cuando existen datos históricos solo se realiza la operación INSERT o UPDATE, según sea el caso, esto viene determinado por la tarea Validación de datos llevada a cabo durante el subproceso de Extracción.
 - ii. Cuando no existen datos históricos se realiza la operación UPDATE. (Kimball & Ross, 2013)

1.7. Pentaho Community Edition

Pentaho es una herramienta de software libre desarrollada por Hitachi, que permite el análisis de BI mediante la conexión con bases de datos relacionales y no relacionales. En su edición de la comunidad (Community Edition), de licencia libre, no posee herramientas propias para la creación de cubos o la creación y almacenamiento de reportes, sin embargo, permite la instalación de herramientas a través de su Marketplace incorporado. Entre las herramientas las más importantes con las que cuenta son:

1.7.1. Schema Workbench

Schema Workbench es una herramienta desarrollada en JAVA para la creación física de cubos OLAP en un formato que Pentaho puede acceder y presentar. Permite, además, la conexión con distintos gestores de Bases de datos para poder obtener de ellos el modelo de la base de datos y generar el cubo de manera más precisa.

1.7.2. Saiku Analytics

Saiku Analytics es una herramienta desarrollada por Metorite BI específicamente para el uso de Pentaho CE, esta herramienta permite la creación y visualización de consultas de manera estructural (tabla) o grafica mediante la implementación de gráficos de barras, de pastel, de dona, histogramas, etc.

Esta herramienta también permite la modificación y creación de nuevas consultas y las almacena dentro del servidor BI lo cual permite su fácil acceso para el análisis respectivo por parte de los usuarios registrados.

1.8. MySQL

MySQL es un gestor de base de datos relacional de código abierto desarrollado comunitariamente, pero adquirido y distribuido por Oracle, posee varias herramientas graficas para su administración, pero la más conocida es phpMyAdmin la cual ha sido desarrollada y distribuida por Apache.

1.9. JSF

Java Server Faces proporciona una arquitectura Modelo-Vista-Controlador muy estructurada que permite manejar la comunicación entre módulos de manera ordenada y segura. Esta tecnología implementa etiquetas con paginas definidas de extensión .xhtml con ayuda de su framework facelets. JSF parte del modelo ya implementado en

JSP permitiendo trabajar con una comunicación entre la Vista (Módulo web) y el controlador separando sus módulos, pero con un constante intercambio de datos.

1.10. Tesseract OCR

Constituye un motor OCR de código abierto que permite la lectura de imágenes mediante el uso de documentos de entrenamiento previamente elaborados. Tiene la capacidad de detectar más de 100 lenguajes y la calidad de sus respuestas dependen mucho del nivel de detalle que disponga su motor de entrenamiento.

Esta tecnología puede ser combinada con lectura y escritura de archivos planos, conversión de formatos como pdf, jpg, png, etc para lograr extraer caracteres en el idioma planteado.

1.11. Apache POI

Elaborado por Apache Software Foundation, es un conjunto de librerías en java que permiten la creación, lectura y escritura de archivos de formato Microsoft Office, utilizadas para realizar extracciones de datos de archivos tipo .xlsx y .docx.

CAPÍTULO 2

2. Diseño y construcción

2.1. Especificación de requerimientos

Para el diseño y construcción del proyecto se han planteado los siguientes requisitos funcionales y no funcionales que detallan las especificaciones del producto deseado:

Requisitos Funcionales

- Permitir el ingreso de documentación inicial como: inscripción de estudiantes, temas propuestos por los tutores, asignación de lectores de plan de Proyecto, inscripción de estudiantes de curso de actualización. Esta documentación será cargada por parte de la Unidad de Titulación.
- Contar con módulos de administración, gestión y aprobación de la información cargada.
- Permitir el ingreso de los estudiantes inscritos después de haber sido aprobados por la Unidad de Titulación para el uso de sus módulos correspondientes.
- Ingresar por parte de los estudiantes documentación definida como: Plan de Proyecto, Borrador del proyecto, estos archivos en formato .doc. También resoluciones del proceso de titulación y comprobantes de pago, estos serán escaneados y transformados a formato .pdf por parte de los estudiantes antes de ser cargados a la plataforma.
- Habilitar la opción de manejar fechas límites para la carga de los archivos por parte de los estudiantes, esto estará a cargo de la Unida de Titulación.
- Permitir la creación de nuevos usuarios para el ingreso al módulo de administración.

- Crear directorios dentro del servidor definiendo: el periodo del proceso de Titulación, si el estudiante se encuentra en cursos de actualización y creando su carpeta personal definida por sus nombres y apellidos que alojará toda la documentación de su proceso de titulación.
- Extraer información de las resoluciones ingresadas por los estudiantes, transforma la misma y la envía al data mart para la realización de reportes. Este proceso lo realiza en tiempo real dando una respuesta al usuario sobre su documento.
- Permitir la visualización de información en módulos de gestión tales como: estudiantes inscritos, estudiantes de actualización y temas propuestos por tutores.

Requisitos no funcionales:

- El software estará alojado en una máquina virtual dispuesta por el departamento del Data Center ICC - Campus Sur.
- La disponibilidad del software está definida por el repositorio en donde se encontrará alojado el software.
- Se dispondrá de IP's definidas para el ingreso al módulo ETL y al módulo de Pentaho para el acceso a los reportes.
- El software contará con métodos estructurados que no excedan un tiempo límite para la carga de los datos.

2.2. Descripción de la solución

En la actualidad la Unidad de Titulación de la carrera de Ingeniería de Sistemas de la Universidad Politécnica Salesiana no cuenta con herramientas definidas que le permitan un análisis rápido y cuantitativo de los indicadores específicos para la toma de decisiones y la oportuna gestión del proceso de titulación de los estudiantes.

El proceso de titulación se encuentra en su totalidad definido, basándose en estatutos y normativas propias de la institución, las tareas que integran el mismo en su mayoría son llevadas a cabo de manera manual y en archivos ofimáticos los cuales no son lo suficientemente eficientes al momento de generar reportes específicos sobre el proceso o de sus tareas.

El proyecto resultante de este análisis, integra herramientas de software que colaboran entre estas, brindando una mejora en la gestión de la información del proceso, la simplificación de las tareas hacia el estudiante y agilidad en la búsqueda de información de acuerdo con indicadores y parámetros relevantes para la Unidad de Titulación.

La solución presentada cuenta con componentes de software comercial en el caso de la base de datos MySQL que permite el almacenamiento de datos con un modelo multidimensional (copo de nieve) y a su vez brinda compatibilidad con Pentaho, esta última descrita como una plataforma para la integración y análisis de datos concordante al paradigma de la inteligencia de negocios (BI).

Para la automatización del proceso se puso énfasis en sus actividades principales, siendo estas:

- Diagramado del proceso a notación BPMN que refleja el flujo de datos y la secuencia de tareas englobadas en el proceso. (Anexo 1)
- Definición de tareas involucradas en el proceso a la vez de los roles que interactúan en el mismo.
- Diseño de módulos web de acuerdo con los roles participantes dentro del proceso.

El módulo de inteligencia de negocios cubre con las siguientes actividades:

- Creación de plantillas con parámetros definidos para el proceso de extracción de datos por parte del ETL.
- Desarrollo de herramienta ETL con entrenamiento de datos basados en la estructura de la documentación manejada por la Unidad de Titulación.
- Modelado del esquema de base de datos que manejará la solución.
- Digitalización de los datos históricos del proceso de titulación para ser mostrados en la herramienta de visualización de información.

2.3. Descripción de los actores

Para la creación de la solución de inteligencia de negocios se analizaron los diferentes participantes dentro del proceso de titulación descritos en las siguientes tablas:

- Unidad de Titulación.

Tabla 2: Actor Unidad de Titulación.

Unidad de Titulación
Descripción
Ente encargado de organizar, planificar, controlar el proceso de titulación de los estudiantes que cumplan con los requisitos para obtener el título de tercer nivel.
Descripción de participación
Participará en todo el proceso de Titulación del estudiante, controlará el mismo mediante el ingreso y extracción de datos además de generar, visualizar reportes del módulo de la solución de BI,

Nota: Tabla que describe al actor Unidad de Titulación.

- Tesorería de sede.

Tabla 3: Actor Tesorería de sede.

Tesorería de sede.
Descripción
Ente encargado dentro de la Universidad Politécnica Salesiana de gestionar las operaciones monetarias del campus.
Descripción de participación
Genera los comprobantes electrónicos que el estudiante cargara a la solución como respaldo del pago efectuado por la gestión en que se encuentra.

Nota: Tabla que describe al actor Tesorería de sede.

- Secretaría de sede.

Tabla 4: Actor Secretaría de sede.

Secretaría de sede
Descripción
Encargada de gestionar los tramites internos dentro del campus, genera, controla y emite documentación de diferentes tipos de proceso.
Descripción de participación
Genera las resoluciones con las que el alumno sigue su proceso de Titulación, maneja los formatos de los documentos que son ingresados a la solución.

Nota: Tabla que describe al actor Secretaría de sede.

- Consejo de Carrera.

Tabla 5: Actor Consejo de Carrera.

Consejo de Carrera
Descripción
Entidad que gestiona a las carreras ofertadas por la Universidad Politécnica Salesiana, así como determinar los diferentes tipos de trabajo de titulación en base a lo establecido por la Unidad de Titulación según el Reglamento de Régimen Académico.
Descripción de participación
Encargada de la aprobación de los documentos que se transformaran en resoluciones para el ingreso en la solución.

Nota: Tabla que describe al actor Consejo de Carrera.

- Dirección de Carrera.

Tabla 6: Actor Dirección de Carrera.

Dirección de Carrera
Descripción
Ente que controla los procesos de la carrea de Ingeniería de Sistemas y Ciencias de la computación.
Descripción de participación
Entidad que recibe y gestiona las peticiones en el proceso de titulación, redirecciona las mismas para ser tratadas en Consejo de Carrera.

Nota: Tabla que describe al actor Dirección de Carrera.

- Estudiante (Inscritos/Curso de actualización).

Tabla 7: Actor Estudiante.

Estudiante
Descripción
Persona que adquiere los conocimientos dentro de la institución y se encuentra culminando o ha culminado su malla curricular.
Descripción de participación
Los estudiantes tienen ingreso a los módulos de la solución en donde podrán cargar su documentación sobre su proceso de Titulación, esto incluye resoluciones, planes y comprobantes de pago que son la principal entrada para la solución de BI.

Nota: Tabla que describe al actor estudiante.

De los actores descritos cabe recalcar que únicamente la Unidad de Titulación es el ente que tendrá acceso a la solución de negocio (ETL y Pentaho). Los departamentos restantes participan dentro del proceso de manera indirecta, generando documentación específica para cada etapa del proceso en el que se encuentre el estudiante.

2.4. Análisis y diseño del módulo ETL

2.4.1. Extracción

La fase de extracción en el software que se ha desarrollado tiene como función obtener los datos delimitados por la unidad de titulación los cuales son de importancia para los análisis a priori que se realizaran. En esta etapa se diferencian dos tipos de archivos de los cuales se obtendrá la información requerida para ser transformada y almacenada en la base de datos multidimensional, así se tiene:

2.4.1.1. Documento en formato pdf

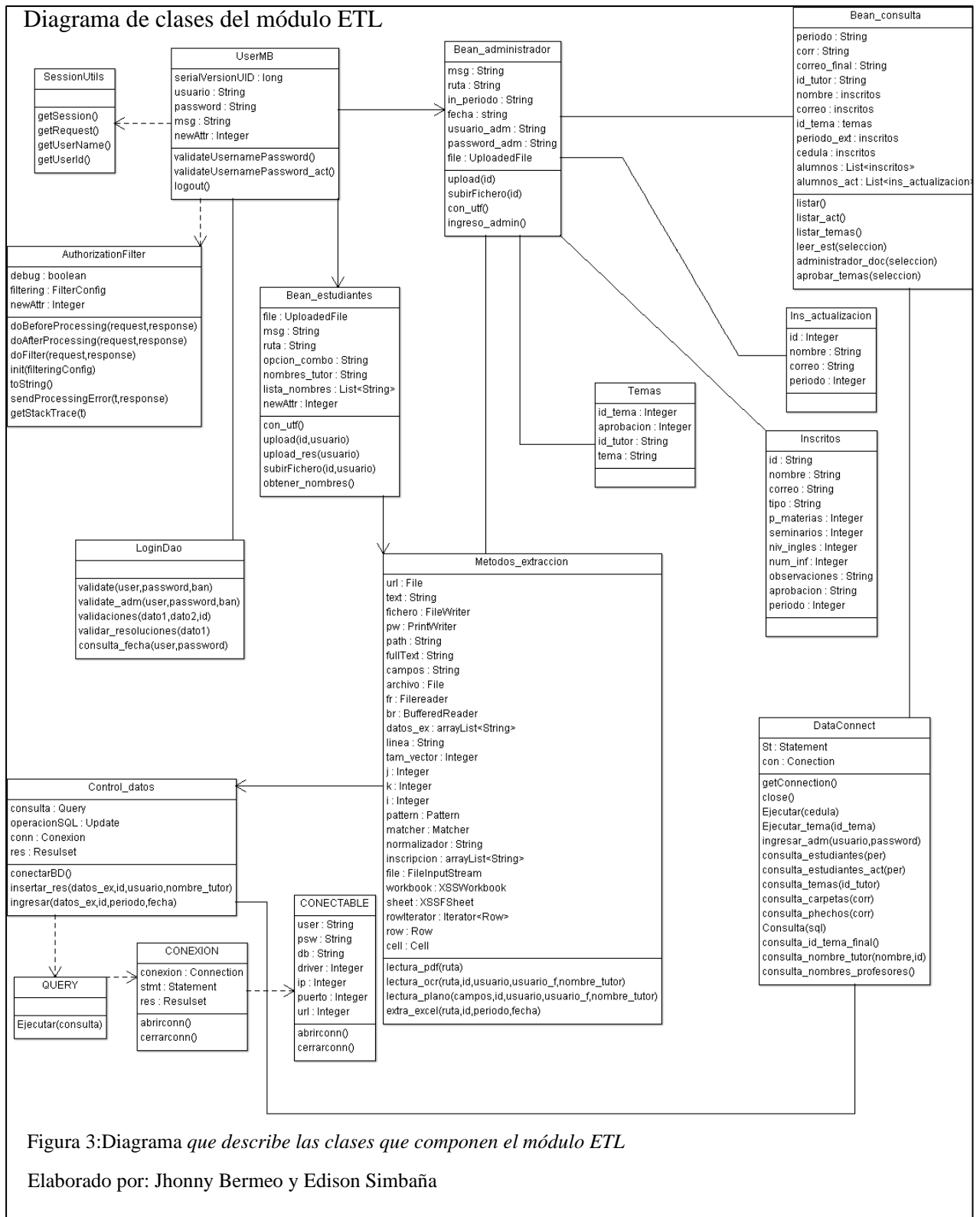
En su mayoría estos serán subidos a la plataforma por los estudiantes que van cumpliendo cada etapa del proceso de titulación. Dentro de estos se pueden diferenciar los archivos de texto que son de uso común y documentos con imágenes dentro de su estructura, estos últimos son analizados por la tecnología ocr para su conversión a texto a continuación seleccionando los parámetros relevantes dependiendo el documento que el estudiante ha subido a la plataforma.

2.4.1.2. Documentos en formato .xlsx (Excel)

La solución cuenta también con la lectura de documentos en este formato. Este se ve utilizado en la lectura de alumnos inscritos al proceso de titulación en donde se seleccionan los primeros datos relevantes de cada estudiante para su posible ingreso dentro del proceso de Titulación. Este paso se lo ve ejecutado por el Coordinador de la unidad de Titulación quien se encargará de cumplir con el formato establecido para que la lectura se ejecute de manera exitosa.

2.4.2. Diseño del ETL

Para el diseño de la herramienta ETL se tomó en cuenta los parámetros necesarios de cada actor que interactúa con el sistema y alimenta el data mart siendo definidas las siguientes clases, variables, constantes y métodos, ver Figura3.



2.5. Análisis y diseño del data mart

Con base en los requerimientos del negocio se diseñó un data mart siguiendo la metodología de Kimball que propone desarrollarlos como entes individuales para posteriormente unificarlos en un único data warehouse, siendo este último paso irrelevante pues, como Kimball propone, un data mart es un repositorio departamental, en este caso del departamento de la Dirección de la Unidad de Titulación de Ingeniería de Sistemas sede Quito.

Antes de modelar la base de datos se procedió con el análisis del proceso de titulación establecido (Anexo 1), el cual indica los participantes en el proceso y cuáles son los parámetros para la obtención de los datos primitivos para la alimentación de la base de datos. Con este análisis se estableció la granularidad para luego modelar las tablas a servir como dimensiones y las cuales representan a los participantes del proceso y los parámetros que darán lugar a los hechos.

El nivel de granularidad para el data mart es: número de alumnos por cada estado del proceso en determinado periodo y por cada opción de titulación.

Tomando en cuenta lo anterior se establecieron los siguientes parámetros para el modelado y construcción del data mart:

- La información relevante de los alumnos se aloja en una única tabla.
- Las opciones de titulación se alojan en una única tabla, debido a que en el futuro podría incrementarse o eliminarse o modificarse alguna opción.
- Las resoluciones y su tipo comparten una sola tabla.
- Para evitar redundancia y sobre escritura de datos de docentes, estos fueron separados en 3 tablas: tutores, lectores y lectores de borrador
- Se estableció una tabla para indicar el estado del proceso al que pertenece la resolución.

- La tabla tema se relaciona directamente con la tabla tutor, sin embargo, debido a la naturaleza del modelo propuesto, esta se relaciona con el tutor mediante la tabla de hechos.
- Se estableció la tabla periodo que actuara como tabla dimensional “tiempo” puesto que indica el momento de ocurrencia de los hechos.
- Siguiendo la metodología adoptada, la tabla de hechos está compuesta por todas las claves foráneas compuestas al relacionar las tablas dimensionales con la tabla central.

Se establecieron los parámetros y factores en base al Proceso de Titulación modelado en BPMN en el Anexo 1, del cual se genera la siguiente matriz proceso/dimensión contenida en la Tabla 8 y definida por la metodología:

Tabla 8: Matriz Proceso/Dimensión para el Proceso de Titulación.

	Dimensiones					
Estados del Proceso de Titulación	Periodo	Opción	Alumno	Lector	Tutor	Resolución
Inicio	X	X	X			X
Gestión	X	X	X	X	X	X
Finalización	X	X	X	X	X	X
Actualización	X	X	X			X

Nota: Descripción del proceso junto a sus dimensiones.

Como se observa en la Tabla No. 8, se identificaron tres estados del Proceso de Titulación y sus respectivas participaciones con las diversas dimensiones propuestas.

Una vez finalizado el análisis se establece que el modelo copo de nieve, establecido al inicio del proyecto y propuesto por Kimball como parte de su metodología para el Datawarehousing, permite la integración de tablas relacionales y su acoplamiento a la tabla de hechos central evitando de esta manera caer en la redundancia en la que cae el modelo estrella.

Las tablas, que describen a los involucrados y a los parámetros se encuentran detalladas con los nombres dentro de la base de datos acogiéndose a los estándares sin el uso de caracteres especiales ni signos, estas intervienen durante el proceso con la siguiente estructura planteada en la Tabla 9:

Tabla 9: Involucrados y tablas.

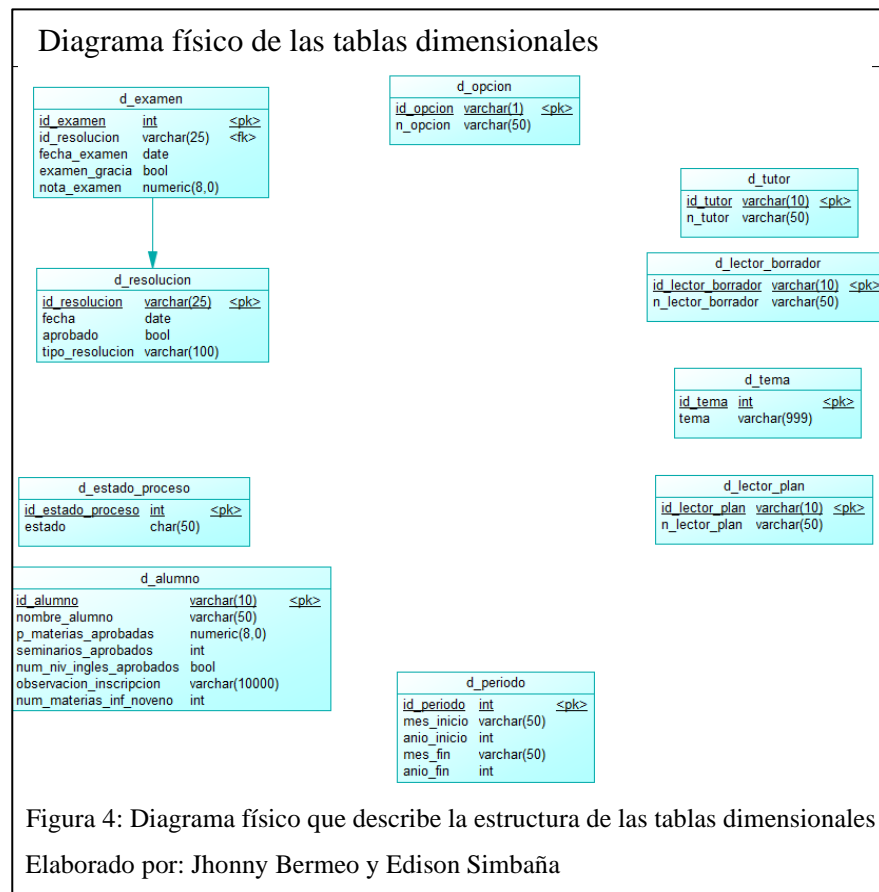
Involucrado/Parametro	Tabla
Alumnos	d_alumno
Docente	d_tutor
	d_lector_plan
	d_lector_borrador
Resoluciones	d_resolucion
Estado del proceso	d_estado_proceso
Temas	d_tema
Opción de titulación	d_opcion
Periodo	d_periodo

Nota: Tabla que describe a los involucrados y sus dimensiones.

Una vez realizado el análisis se procede con el modelado, tomando en cuenta que el modelo multidimensional de BDD, según la metodología de Kimball, está compuesto por tres elementos: Dimensiones, hechos y medidas.

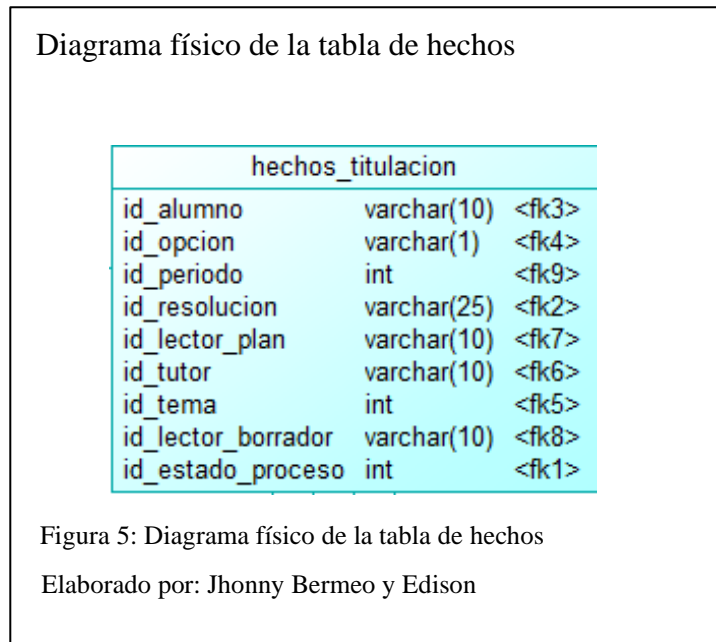
2.5.1. Dimensiones

Las dimensiones, o tablas dimensionales, son aquellas que contienen la información descriptiva de los entes y parámetros que intervienen en el proceso de titulación de la UT. Surgen a partir del análisis que se realiza en base a la granularidad propuesta y dan la pauta para la tabla de hechos. Cada tabla describe un miembro o un parámetro relevante en el proceso de titulación y poseen un identificador para cada elemento (clave primaria). El modelo de las tablas dimensionales obedece a la Figura 4:



2.5.2. Tabla de hechos

Al concluir el análisis para las dimensiones se procede con el análisis y modelado de la tabla de hechos, esta tabla presenta la estructura acorde a los requerimientos del negocio previstos en la dimensión. El modelo de la tabla de hechos está compuesto por los identificadores de las dimensiones y tiene la estructura descrita en la Figura 5.



Una vez modeladas las tablas dimensionales y la tabla de hechos se procede con el siguiente paso siguiendo la metodología Kimball, la cual consiste en el levantamiento físico de la base de datos. Para esto se utilizó la herramienta PowerDesigner la cual genera automáticamente el script de la base de datos para ser implementado en MySQL, de esta manera se obtiene la estructura de la Figura 6:

Diagrama físico de la estructura copo de nieve

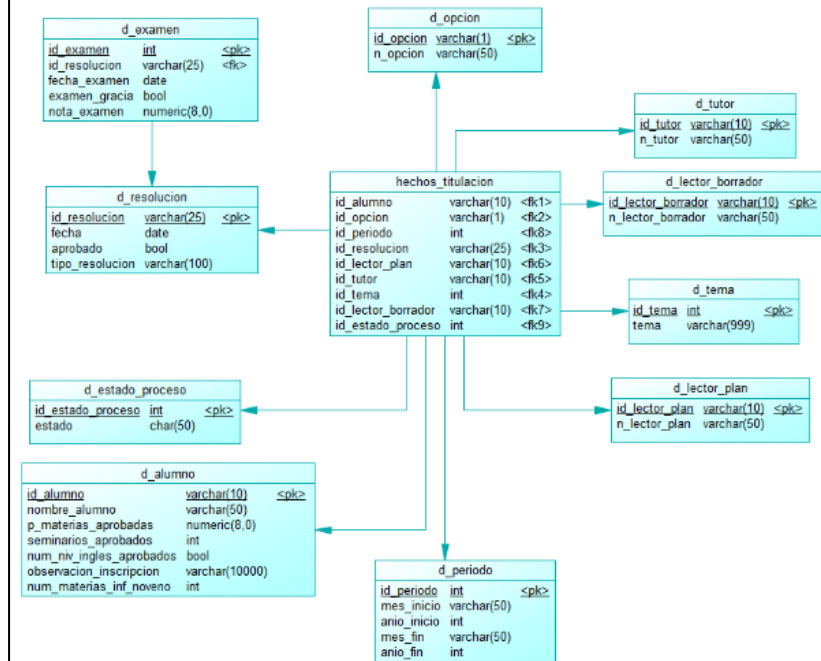


Figura 6: Diagrama físico que muestra la estructura copo de nieve de la base de datos multidimensional

Elaborado por: Jhonny Bermeo y Edison Simbaña

La base de datos de estructura copo de nieve ahora se encuentran en MySQL, sin embargo, Pentaho no puede acceder directamente a la base de datos para realizar el análisis, por ende, se debe de proceder con el levantamiento físico del cubo. Para esto se hizo uso de la herramienta SchemaWorkbench de Pentaho la cual permite la conexión con la base de datos relacional (implementada en MySQL) obteniendo de esta manera las tablas que formaran parte del cubo.

2.5.3. Medidas

Para establecer las medidas se debe tomar en cuenta que estas solo pueden provenir de la tabla de hechos, de modo que, en este caso, son el resultado del conteo de parámetros que resultan de interés tanto para la dirección de la UT como para autoridades de la UPS.

Se procede con el análisis de este requerimiento, tomando en cuenta que la base de datos primaria que maneja la Unidad de Titulación tiene como principal elemento al número de alumnos inscritos, en cada etapa del proceso, por cada periodo académico. Con este antecedente se puede determinar que la medida principal para el análisis respectivo es el número de alumnos en los distintos parámetros establecidos del proceso de titulación. La medida relevante es, por tanto, el conteo de alumnos presentes o inscritos en el Proceso de Titulación desde el periodo académico No. 46 hasta el periodo actual (55).

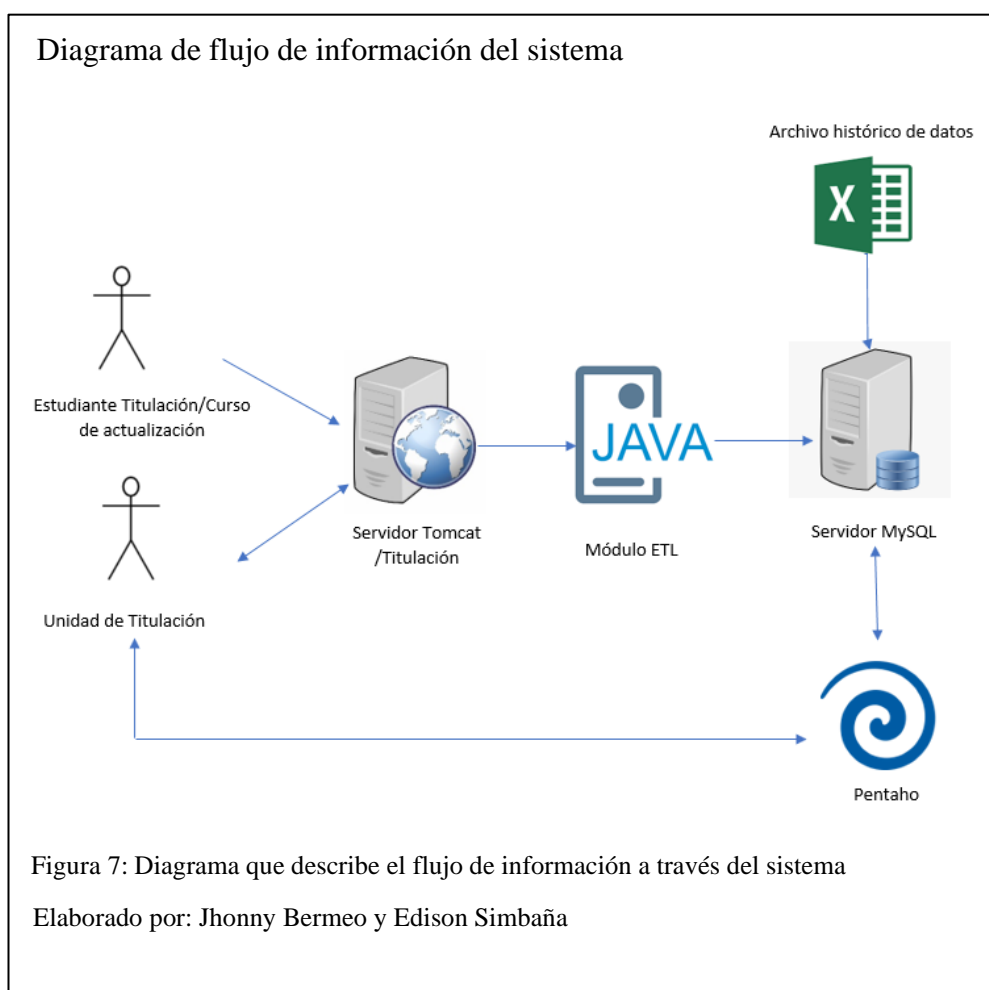
2.5.4. Análisis ROLAP

Una vez que los parámetros, dimensiones, tabla de hechos y medidas se han determinado, se procede con el levantamiento de la herramienta para realizar el análisis de datos en línea de la base de datos relacional (ROLAP). La herramienta seleccionada es una herramienta propia de la suite de Pentaho Community Edition, Saiku Analytics, la cual es de licencia de uso libre y provee de un entorno sencillo para realizar las tareas necesarias para la creación de reportes y el posterior análisis.

CAPÍTULO 3

3. Explotación

El proceso de Titulación describe un flujo de información por los módulos especificados en la Figura 7.



Ingreso inicial

Se basó en un archivo histórico que contiene los datos de la Unidad de Titulación hasta el periodo 54, estos fueron cargados de manera manual al data mart.

Ingreso Periódico

De manera paulatina los datos desde el periodo 56 y parte del periodo 55 se ingresarán mediante la herramienta ETL donde se verificará su validez y se cargará en el data mart.

3.1. Descripción del proceso de ingreso de datos al ETL

3.1.1. Módulo Inscripción

El proceso de Titulación se inicia con la apertura de inscripciones por parte de la dirección de la unidad de Titulación, para este objetivo se realiza la petición de los datos iniciales sobre los estudiantes para realizar una verificación y valides de los mismos. Debido a este antecedente se toma a los estudiantes que cumplen con los requisitos planteados con la información inicial y se los carga en un archivo .xlsx para ser enviados al data marta desde su respectivo módulo del administrador.

Campos Requeridos:

- **Id del estudiante:** Número de cedula del estudiante.
- **Apellidos y Nombres:** Datos completos del estudiante en ese orden, los signos y la diferenciación entre mayúsculas y minúsculas no afecta la lectura.
- **Correo Institucional:** Correo proporcionado de la institución que se convertirá en el id de usuario futuro del estudiante.
- **Opción de Titulación:** **P** para proyecto técnico, **A** para artículo académico y **E** para examen complejo, de igual manera el tipo de letra mayúscula o minúscula no afecta su lectura.

- **Porcentaje de materias aprobadas:** Porcentaje extraído de la revisión de la documentación enviada por el estudiante a cargo de la Unidad de Titulación.
- **Número de seminarios aprobados:** Numero de seminarios que constan aprobados por parte del estudiante.
- **Número de niveles de inglés aprobados:** Niveles de inglés aprobados por parte del estudiante hasta la fecha de inscripción.
- **Número de materias inferiores a noveno:** Cantidad de materias que el estudiante no ha cursado hasta la inscripción.
- **Observaciones:** Si existen detalles que deban ser tomados en cuenta acerca de los datos proporcionado por el estudiante.
- **Periodo de inscripción:** Periodo en el que se realiza la inscripción del estudiante, será común en todos los registros a menos que se planteen casos especiales.
- **Estado de aprobación:** Indicador que muestra el estado de la inscripción del estudiante antes de la aprobación final por parte de la unidad de Titulación, la misma que se realizara en el módulo de gestión de estudiantes después de cargar esta información.

Formato de ingreso de Inscritos a UT

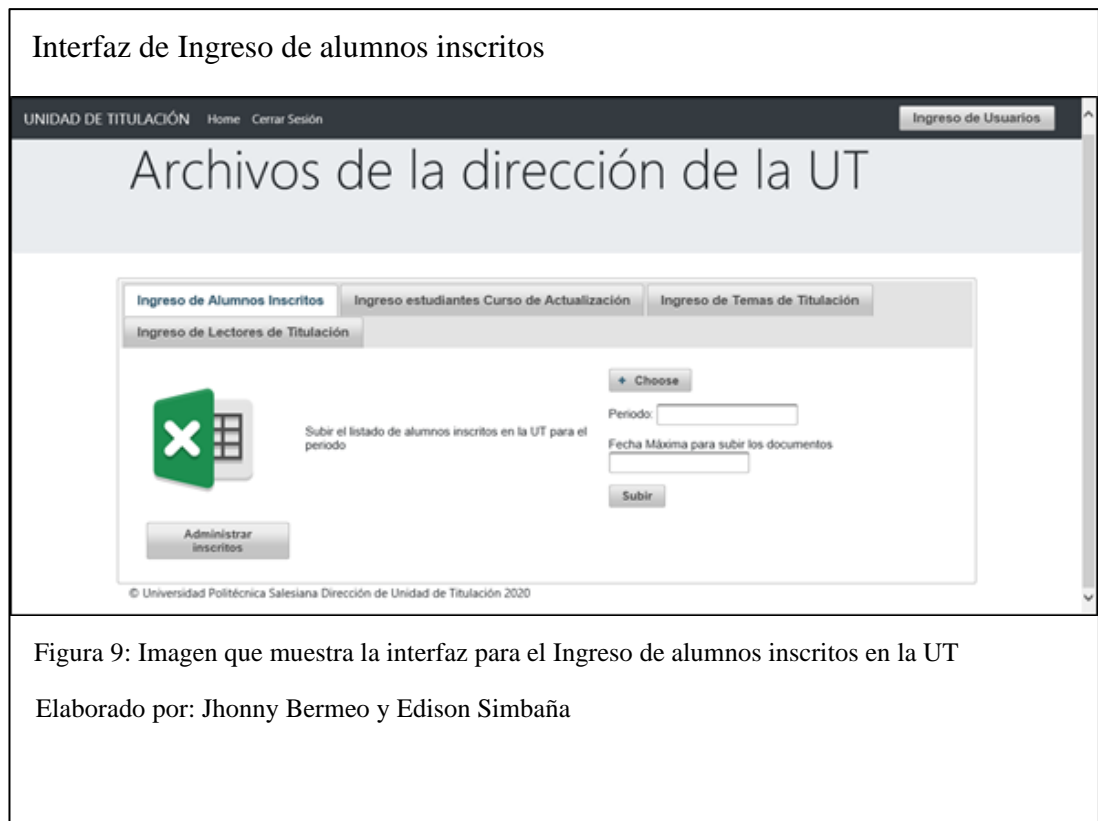
M2	APELLIDOS Y NOMBRES		OPCIÓN DE TITULACIÓN	# DE SEMINARIOS APROBADOS	# DE MATERIAS INFERIORES A NOVENO	PERIODO DE INSCRIPCIÓN			
	A	B	C	E	G	I	K		
1	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec	100	5	6	2	55 negado	
2	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec	A	90	4	6	5	55 negado
3	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec	E	80	5	5	1	55 negado
4	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec	P	50	4	6	3	55 negado
5	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec		100	5	5	6	No se define la opción de titulación. 55 negado
6	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec		80	4	4	4	55 negado
7	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec	P	90	4	5	2	55 negado
8	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec		70	5	6	3	Inglés IV cursando, Inglés V no cursado 55 negado
9	XXXXXXXXXX	XXXX XXXX XXXXXXXX XXXXXXXXXX	estudiante@ups.edu.ec	E	80	4	4	1	55 negado
10									
11	ID DEL ESTUDIANTE		CORREO INSTITUCIONAL	PORCENTAJE DE MATERIAS APROBADAS	# DE NIVELES DE INGLES APROBADOS	OBSERVACIONES	ESTADO DE APROBACIÓN		
12									

Figura 8: Imagen que muestra el formato del archivo para el ingreso de Inscritos a UT
Elaborado por: Jhonny Bermeo y Edison Simbaña

Como se presenta en la Figura 8, no es necesario ningún tipo de encabezado para la lectura de los datos en el archivo .xlsx.

Para la carga de este archivo el coordinador de la Unidad de Titulación después de realizar el logeo con las credenciales autorizadas se dirigirá a la opción de Ingreso de alumnos inscritos representada en la Figura 9, donde junto al archivo antes estructurado se ingresarán:

- **Periodo:** Periodo en el que el Coordinado de la Unidad de Titulación realiza la inscripción de los estudiantes.
- **Fecha:** Fecha máxima para que los estudiantes ingresen todos los documentos correspondientes a su proceso de titulación. Cabe destacar que el alumno no podrá ingresar documentos si la fecha que ha descrito el coordinador de la Unidad de Titulación ha expirado.



3.1.2. Módulo Curso de Actualización

Los estudiantes que ingresan al curso de actualización cuentan con un módulo diferente al de los estudiantes en donde deben subir la documentación correspondiente a su estado dentro de la institución, para este motivo el Coordinador de la unidad de Titulación ingresará al sistema un documento con la información de los estudiantes que cuenta con los siguientes datos:

- **Id del estudiante:** Número de cédula del estudiante que será utilizado como contraseña.
- **Apellidos y Nombres:** De la misma manera deben constar los apellidos y nombres completos en ese orden.
- **Correo Institucional:** Correo proporcionado a los estudiantes de actualización para tomar las materias del curso.

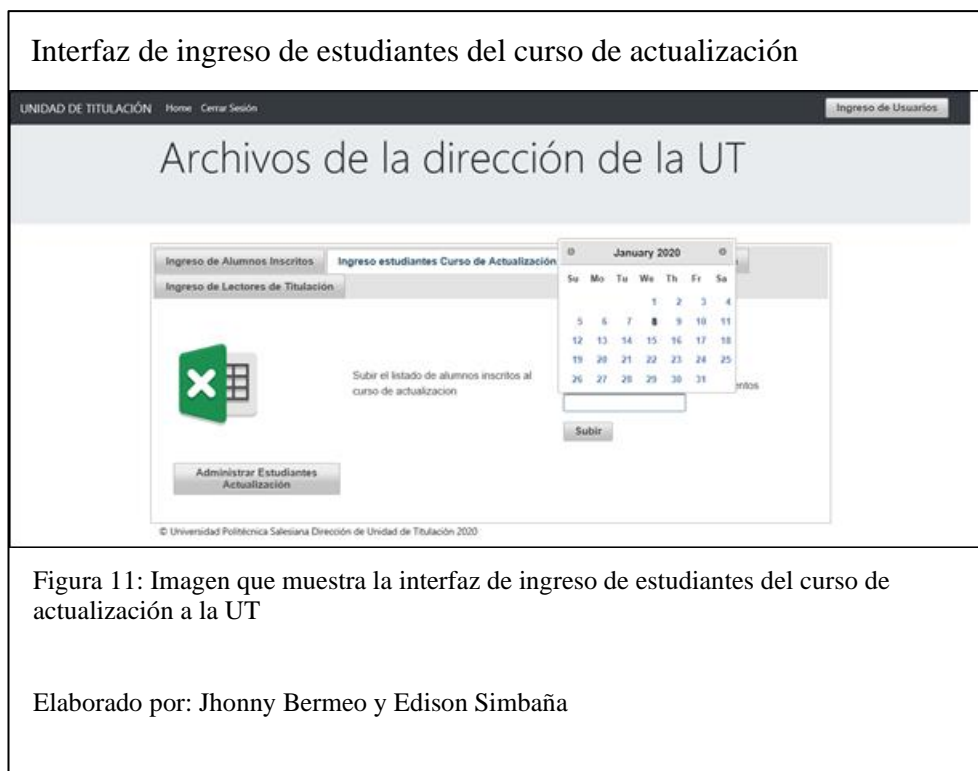
Formato de ingreso estudiantes del curso de actualización

B17				
APELIDOS Y NOMBRES				
	A	B	C	D
1	XXXXXXXXXX	XXXXXX XXXXXX XXXXXXXX XXXXXXXX	estudiante@ups.edu.ec	
2	XXXXXXXXXX	XXXXXX XXXXXX XXXXXXXX XXXXXXXX	estudiante@ups.edu.ec	
3	XXXXXXXXXX	XXXXXX XXXXXX XXXXXXXX XXXXXXXX	estudiante@ups.edu.ec	
4				
5				
6				
	ID DE ESTUDIANTE		CORREO ELECTRÓNICO	

Figura 10: Imagen que muestra el formato del archivo para el ingreso estudiantes del curso de actualización a la UT

Elaborado por: Jhonny Bermeo y Edison Simbaña

De similar modo para cargar los datos de los estudiantes del curso de actualización se lo realiza sin ningún tipo de encabezado como se presenta en la Figura 10, el coordinador de la Unidad de Titulación asigna el periodo de inscripción de los estudiantes y también la fecha máxima para que el estudiante suba su documentación, estos datos son ingresados en la interfaz representada en la Figura 11.



3.1.3. Ingreso de Temas de Titulación

El ingreso de temas para cada periodo se determina entre el docente tutor que propone el tema y la Unidad de Titulación, los temas ingresados en este apartado serán evaluados por el Consejo de Carrera y posteriormente aprobados para ser publicados ante los estudiantes inscritos al proceso.

Son necesarios en el archivo los siguientes puntos delimitados en la Figura 12:

- **ID del tutor:** Número de cédula del docente tutor que propone el tema.
- **Apellidos y nombres:** Apellidos y nombres completos del tutor, este dato es de mucha relevancia para comparar dentro de los docentes que constan en el data mart y en caso de no existir serán ingresados de manera automática.
- **Tema:** Tema propuesto por el docente tutor.

Formato de ingreso de temas de proyectos

J12 ID TUTOR				TEMA DE PROYECTO TÉCNICO O ARTÍCULO ACADÉMICO			
A		B		C		D	
1	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	Desarrollo de modulo gerencial integrador financiero en el sistema RUA, para las Obras y Casas Salesianas en el Ecuador.				
2	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	Desarrollo de un Sistema de Administración de biométricos para las unidades del distrito 17d04.				
3	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	Desarrollo de una aplicación para la creación de matrices y manejo estadístico del Distrito 17D04 y unidades de salud.				
4	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	Desarrollo un Sistema de agendas de pacientes para las unidades del Distrito 17D04				
5	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	Diseño de un sistema e-commerce para los emprendimientos de Cayambe				
6	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	Sistema de gestión de horarios y asignación de recursos para el DataCenter de la Carrera de Ingeniería en Ciencias de la Computación				
7	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	DISEÑO E IMPLEMENTACIÓN DE UNA RED WAN CORPORATIVA PARA EL DISTRITO 17D04 "CENTRO DE SALUD N.º 1 CENTRO HISTÓRICO"				
8	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	DISEÑO E IMPLEMENTACIÓN DE UNA APLICACIÓN ADAPTIVA (SERVIDOR DE APLICACIONES) PARA CONTROL DOMOTICO DENTRO DEL ESTANDAR ISO/IEC 25010 BASADOS EN IoT.				
9	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	DISEÑO E IMPLEMENTACIÓN DE UN PROTOTIPO DE UNA ARQUITECTURA DE MICRO SERVICIOS APLICADA EN CAPA MEDIA MIDDLEWARE PARA APLICACIONES WEB.				
10	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	DISEÑO E IMPLEMENTACIÓN DE CONTROL DOMOTICO (CON SISTEMAS EMBEBIDOS) PARA CONECTARSE CON APLICACIONES ADAPTIVAS BASADOS EN IoT.				
11	XXXXXXXXXX	XXXXX XXXXX XXXXXXX XXXXXXXX	DISEÑO E IMPLEMENTACIÓN DE UNA APLICACIÓN MÓVIL PARA LA ESCUELA FISCOMISIONAL "SOLIDARIDAD" DENTRO DEL ESTANDAR ISO/IEC 25010 QUE PERMITA EL SEGUIMIENTO ACADEMICO DE LOS ESTUDIANTES POR PARTE DE LOS PADRES DE FAMILIA.				
		APellidos y Nombres Tutor	Desarrollar un sistema que permita la comunicación directa entre un médico y cualquier persona presente en el lugar de emergencia, con el fin de brindar				

Hoia1

Figura 12: Formato de ingreso de temas de proyectos

Elaborado por: Jhonny Bermeo y Edison Simbaña

El documento no contiene encabezado de la misma manera que los archivos anteriormente descritos. Para este apartado no es necesario el ingreso de otros parámetros además del documento planteado.

Administración de Temas

Los temas pueden ser administrador con el botón **Administrar Temas** donde se redirigirá al Coordinador de Unidad de Titulación a la interfaz específica para este objetivo.

La administración de temas tiene lugar cuando los temas han sido evaluados y aprobados o negados en consejo de carrea, de esta manera se podrá cuantificar la eficacia de los temas y docentes que serán presentados en la herramienta de reportes.

3.1.4. Ingreso de lectores de planes de titulación

El ingreso de lectores de planes de titulación tanto para proyecto técnico como para artículo académico se realizará después de que los estudiantes están inscritos y aprobados, además de que los temas han pasado por la evaluación del Consejo de Carrera, para este objetivo se necesitan los siguientes datos también representados en Figura 13:

Tema: Tema de Proyecto Técnico o Artículo Académico aprobado por el Consejo de Carrera.

Id de tema: Identificador de tema ingresado en el módulo de **Ingreso de Temas de Titulación**, este parámetro puede y debe ser consultado desde la administración de temas del módulo especificado anteriormente.

Tipo de opción de titulación: Si el estudiante no ha especificado su tipo de Titulación en la inscripción, en este documento puede ser ingresado este dato.

Id Tutor: Número de cédula del tutor del estudiante.

Apellidos y nombres del tutor: Información completa del tutor para ser comparado en el data mart.

Id lector de plan de titulación: Número de cédula del lector asignado para la lectura del proyecto, de la misma manera será validado la existencia del lector en el data mart y de no estar registrado se lo realizará de manera automática.

Apellidos y nombres lector de plan de titulación: Información completa del lector que servirá también para la comparación con los del data mart.

Id estudiante 1: Id del estudiante que fue ingresado en la inscripción al proceso.

Apellidos y nombres estudiante 1: Datos completos del estudiante 1.

Id estudiante 2: Id del estudiante que fue ingresado en la inscripción al proceso.

Apellidos y nombres estudiante 2: Datos completos del estudiante 2.

Nota: En caso de existir un único estudiante cursando el proceso de Titulación este deberá estar declarado a continuación de los datos del Lector de Plan de Titulación sin dejar espacios en blanco en el documento.

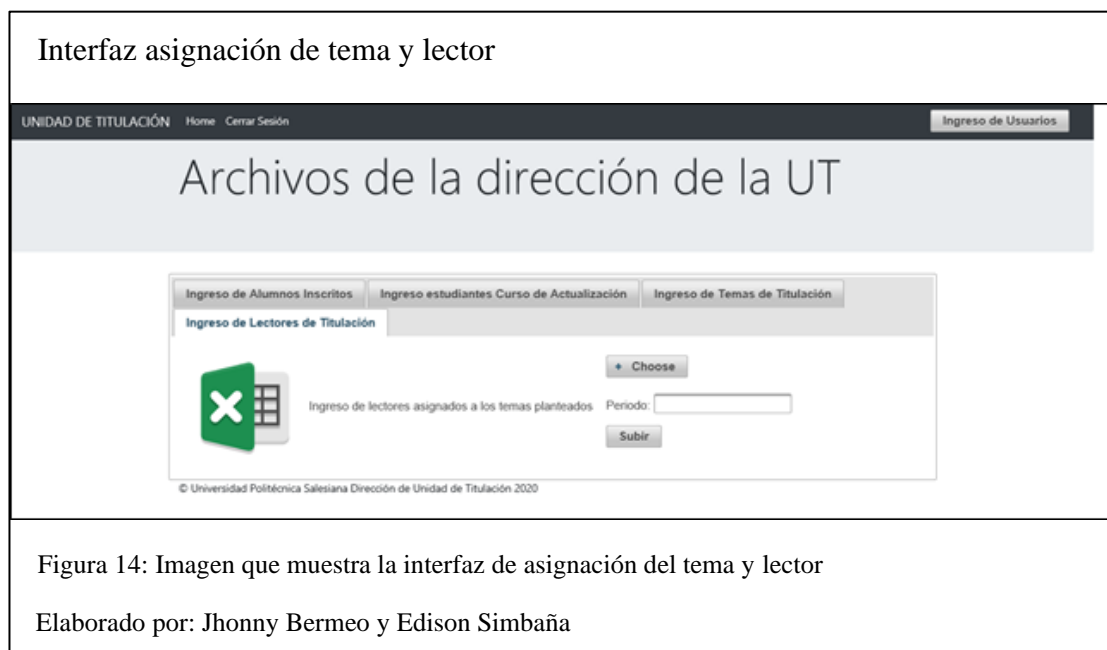
Formato de asignación de tema y lector

	ID DEL TEMA	ID TUTOR	ID LECTOR DE PLAN	ID ESTUDIANTE 1	APellidos y Nombres ESTUDIANTE 1				
1	Desarrollo de modulo gerencial integrador financiero en el sistema RUA, para las Obras y Casas Salesianas en el Ecuador.	237	P	XXXXXXXXXXXX	XXXXXXXXXXXX	XXXXXXXXXXXX	XXXXXXXXXXXX	XXXXXXXXXXXX	XXXXXXXXXXXX
2									
3	TEMA	TIPO DE PROYECTO	APellidos y Nombres TUTOR	APellidos y Nombres LECTOR PLAN	ID ESTUDIANTE 2	APellidos y Nombres ESTUDIANTE 2			
4									

Figura 13: Imagen que muestra el formato del archivo para la asignación de tema y lector

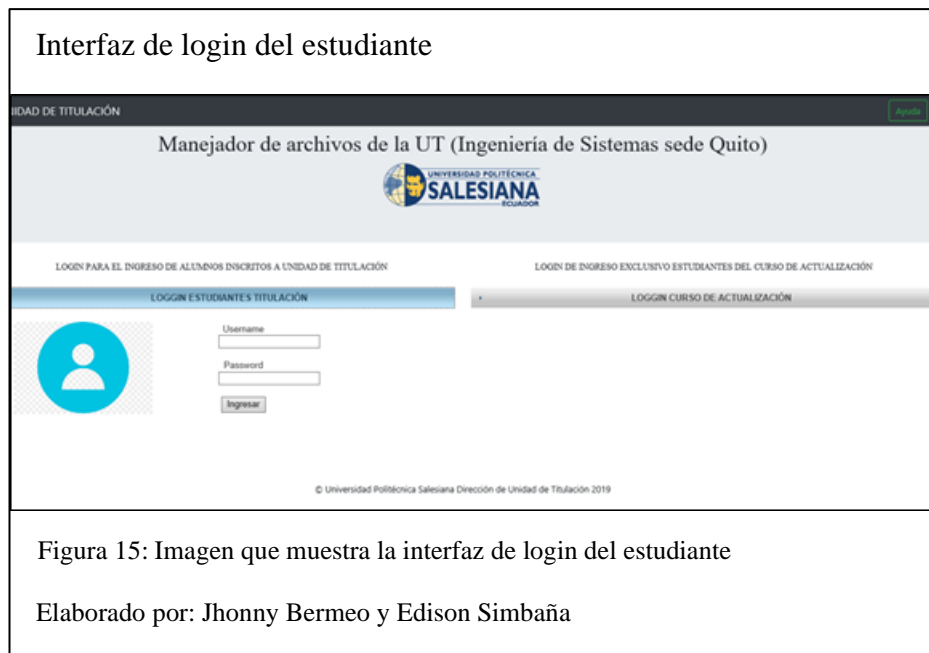
Elaborado por: Jhonny Bermeo y Edison Simbaña

El ingreso de los Lectores se realiza en el módulo **Ingreso de Lectores de Titulación** como se presenta en Figura 14, no es necesario el ingreso de parámetros extra para esta documentación, es de gran importancia que los datos sean revisados antes de ser subidos al sistema.



3.1.5. Ingreso Inicio de Titulación (Estudiantes inscritos y aprobados)

El estudiante aprobado podrá ingresar al sistema con sus credenciales como se presenta en Figura 15, después de haber sido aprobado por la unidad de Titulación y mientras se encuentre dentro de la fecha máxima para la carga de sus documentos.

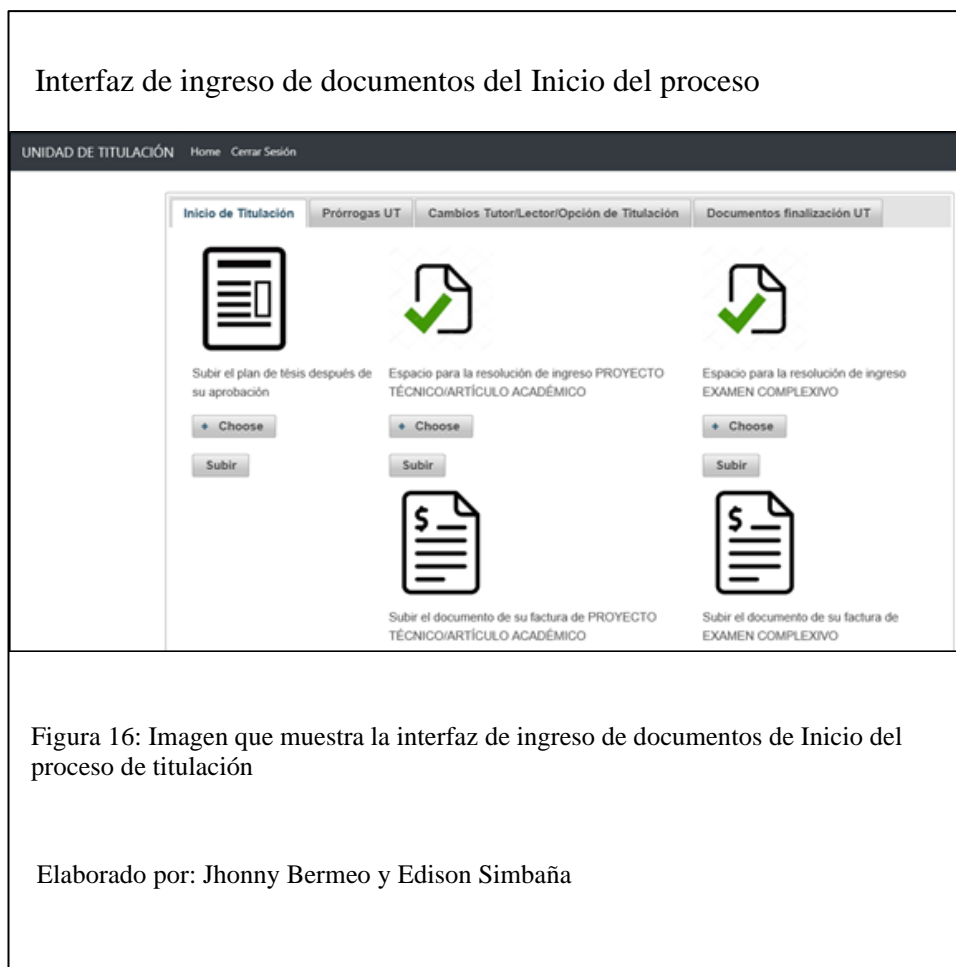


De ser exitoso el logeo el estudiante inscrito a la Unidad de Titulación podrá visualizar la página principal con los módulos, estos se encuentran separados de manera que el usuario pueda entender la navegabilidad del sitio

Inicio Titulación

El módulo representado en la Figura 16 presenta el espacio específico para los documentos que inician el proceso de Titulación:

- **Plan de Proyecto:** Plan realizado por los estudiantes donde se describe el proyecto de forma rápida y ha sido aprobado para su realización.
- **Resolución de Ingreso:** Resolución emitida por secretaria de campus, después de que el proyecto ha sido aprobado por el Consejo de Carrera.
- **Comprobante de pago:** Documento emitido de manera electrónica como respaldo de la cancelación de los valores de inscripción.
- **Resolución de Examen complejo:** Resolución emitida que declara la autorización del estudiante para rendir el examen. El estudiante que opta por esta opción no debe cargar ningún archivo de plan de proyecto.



Prórrogas UT

Módulo dedicado a gestionar tanto prórrogas a las que pueden acceder los estudiantes dentro de su proceso de titulación como también el ingreso de resoluciones de diferente índole descritas a continuación y representadas en Figura 17:

- **Resolución de primera prórroga:** Resolución aprobada que comunica que el estudiante ha realizado su primera prórroga de tiempo dentro de su proceso.
- **Resolución de segunda prórroga:** Resolución aprobada, indica que el estudiante no ha terminado su proyecto y necesita de un segundo periodo para la culminación de este. Esta resolución genera un pago adicional que debe ser cargado, además que dependiendo del inicio del proyecto del estudiante esta podría significar un posible ingreso del estudiante al curso de actualización.

- **Resoluciones Adicionales:**
 - **Ajuste de Plan de Proyecto de Titulación:** El estudiante ha solicitado el cambio de parámetros y detalles dentro de su plan de Proyecto.
 - **Anulación de Proyecto de Titulación:** El estudiante opta por anular su proyecto, al optar por esta opción el estudiante dejara de tener acceso a la plataforma.
 - **Cambio de Tema de Proyecto de Titulación:** Solicitud que indica que el estudiante ha cambiado de tema de proyecto.
 - **Continuidad de proyecto de Titulación:** El estudiante retoma las actividades dentro de su proyecto.
 - **Examen complejo de gracia:** Resolución que indica que se ha aprobado que el estudiante podrá acceder a una segunda oportunidad después de haber fallado en su primer intento en su opción de rendir el examen complejo.
 - **Modificación de Tema de Proyecto de Titulación:** Resolución que aprueba la modificación del tema solicitado por el estudiante.
 - **Renuncia a Proyecto de Titulación-Tutor:** Resolución que comunica la renuncia al cargo de tutor del proyecto solicitada por el tutor, esta resolución termina con el proceso de titulación del estudiante al igual que sus credenciales para el ingreso al módulo.
 - **Renuncia a Proyecto de Titulación-Autor:** En este caso la renuncia al proyecto se produce por parte del autor del proyecto, de la misma manera sus credenciales serán desactivadas para el ingreso a la plataforma.

- **Resolución de tercera prorroga:** Resolución que indica que el consejo de carrea ha permitido una tercera prorroga al estudiante, también podría generar el ingreso del estudiante al curso de actualización.
- **Resolución de no publicación de Proyecto de Titulación:** El estudiante solicita que su proyecto no sea publicado debido a razones particulares.

Todas las resoluciones ingresadas deben ser escaneadas y cargadas al sistema en formato .pdf con la mayor claridad posible.

Si las resoluciones han generado comprobantes de pago estos deberán ser cargados también en el apartado correspondiente.

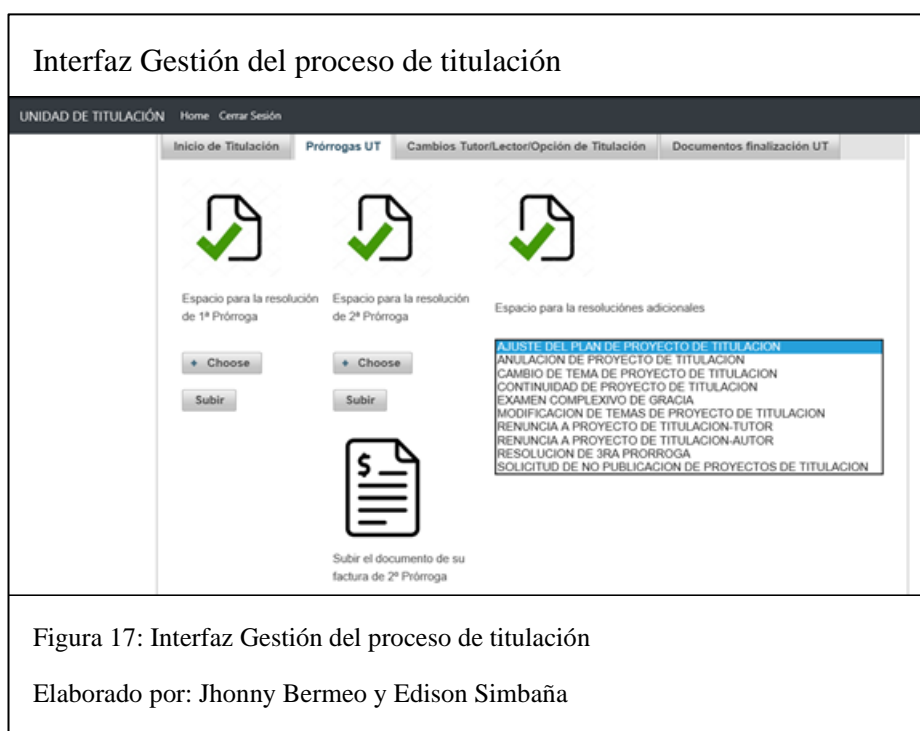


Figura 17: Interfaz Gestión del proceso de titulación

Elaborado por: Jhonny Bermeo y Edison Simbaña

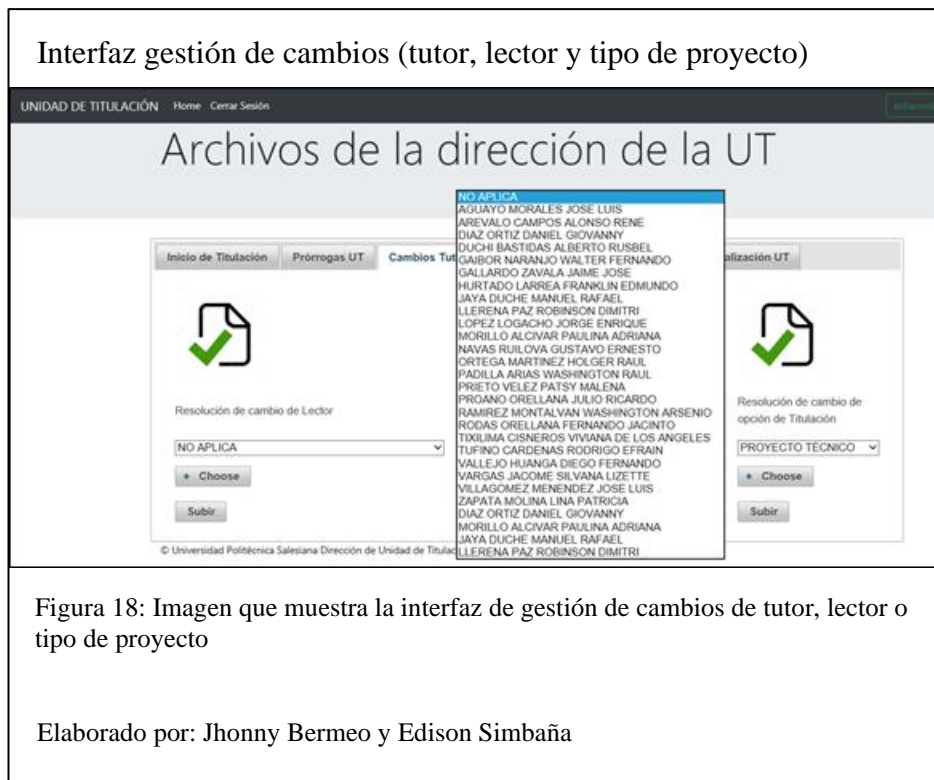
Cambio Tutor/Lector/Opción de Titulación

El módulo de cambios presentado en la Figura 18 permite al estudiante comunicar a la Unidad de titulación acerca de los cambios que se aplica a su proceso, con la finalidad de mantener la integridad de los datos se presenta las opciones de Lector, Tutor y Opción de Titulación disponibles en el data mart, el estudiante elegirá su resolución dependiendo su trámite y también la opción a la que desea acceder.

Si la resolución ha sido aprobada los cambios serán efectuados dentro del data mart.

- **Resolución de cambio de lector:** Resolución que aprueba el cambio de lector del borrador del proyecto del estudiante.
- **Resolución de cambio de tutor:** Resolución que indica el cambio del tutor que continuara con el proyecto del estudiante.
- **Resolución de cambio de opción de Titulación:** Resolución que aprueba el cambio de la opción del estudiante, sea esta Proyecto Técnico, Artículo Académico o Examen Complexivo.

Cabe recalcar que, si el estudiante tenía como antecedente un proyecto Técnico o Artículo académico y accede al Examen Complexivo, deberá regresar al módulo anterior y subir la documentación que avale su nuevo proceso.



Documentos Finalización UT

El módulo de finalización es la etapa que culmina con el proceso de titulación, es subproceso que maneja pocas resoluciones, pero se centra en la documentación que genera la entrega del proyecto, los documentos que no son catalogados como resoluciones en este apartado tales como informes y autorizaciones pueden ser cargados en formato .pdf de preferencia o en .doc si es el caso como se ilustra en la Figura 19.

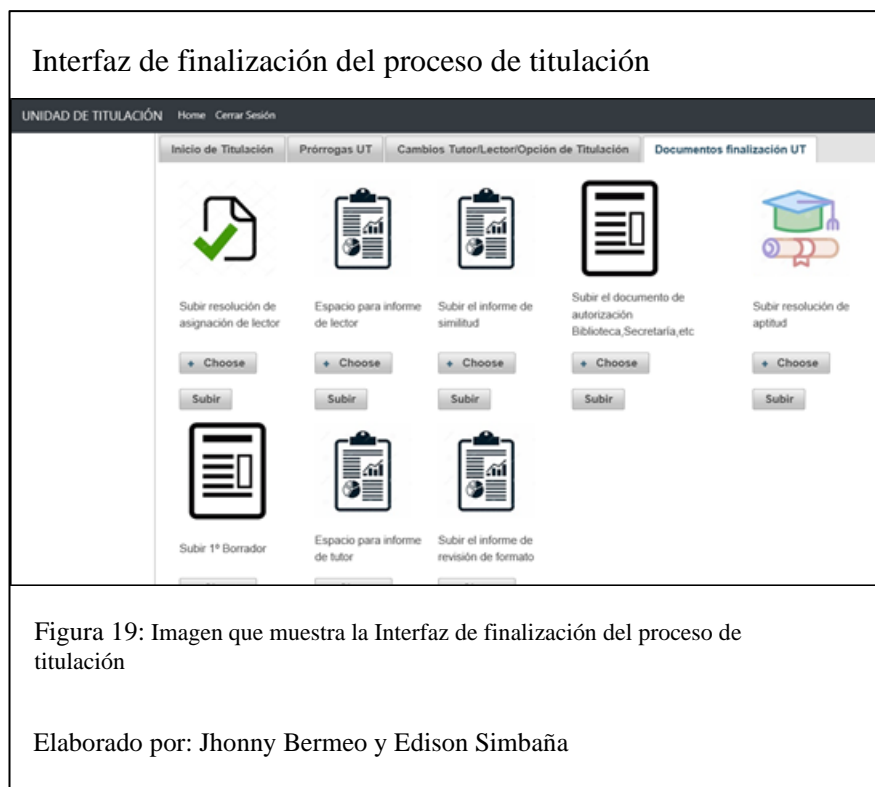


Figura 19: Imagen que muestra la Interfaz de finalización del proceso de titulación

Elaborado por: Jhonny Bermeo y Edison Simbaña

3.1.6. Ingreso estudiantes Curso de Actualización

El estudiante de actualización tiene su espacio de logeo en la parte derecha de la interfaz de inicio como se muestra en la Figura 20, en donde sus credenciales serán verificadas para su ingreso. Existe la posibilidad de que un estudiante del curso de actualización también ingrese al módulo de estudiantes inscritos, en este caso el estudiante podrá ingresar a ambos módulos en simultaneo mientras se encuentre cursando su actualización.

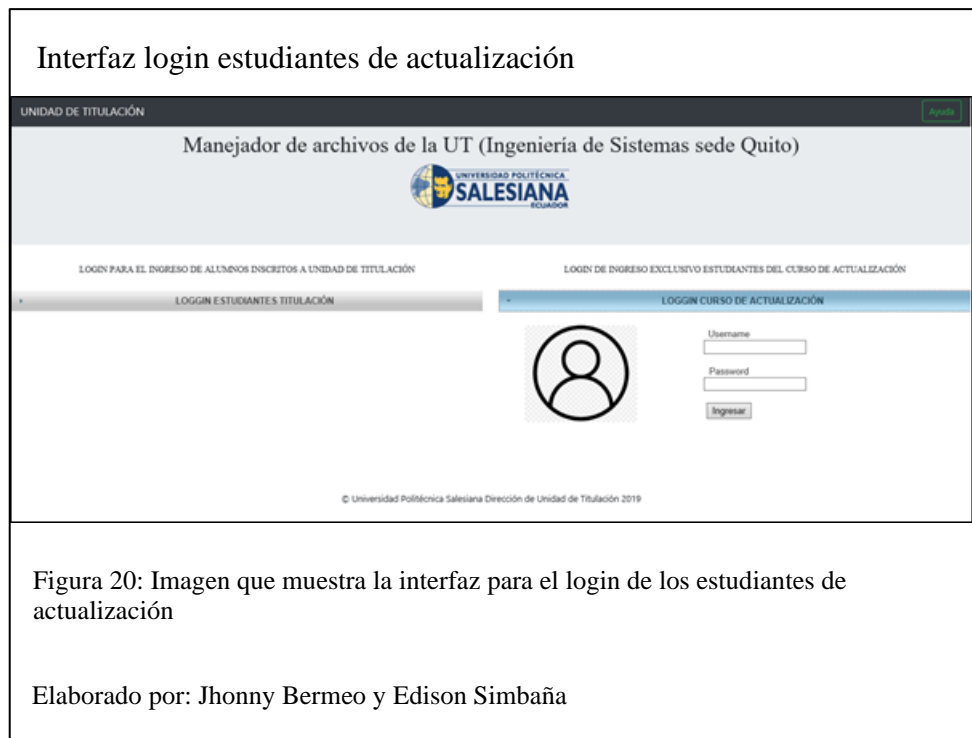


Figura 20: Imagen que muestra la interfaz para el login de los estudiantes de actualización

Elaborado por: Jhonny Bermeo y Edison Simbaña

Al verificar que las credenciales son correctas, el estudiante de actualización tendrá acceso a la interfaz y módulos que especifican los documentos requeridos de este subproceso.

- **Resolución de ingreso al curso de actualización:** Resolución que aprueba el ingreso del estudiante al curso de actualización después de haber sido evaluada en Consejo de Carrera.
- **Comprobante de pago:** Factura electrónica enviada al correo del estudiante, generada después que el estudiante cancela los valores que conlleva el ingreso al curso.
- **Formulario de ingreso:** Documento que contiene los datos del estudiante y la petición de ingreso al curso.

3.2. Descripción del proceso de ingreso de datos al data mart

Para el ingreso de datos, en la base de datos relacional, en primer lugar, se hizo un análisis, de los datos disponibles, para determinar cómo estos deberían ser ingresados a la base de datos.

Los datos pertenecientes al periodo académico 46 hasta el periodo académico 54 (disponibles en la base de datos primaria de la Unidad de Titulación) en la Figura 21, fueron determinados como datos históricos e ingresados de manera manual a la base de datos.

Base de datos primaria de la UT

OPCIÓN DE TITULACIÓN						OPORTUNIDAD																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO	EXAMEN COMPLETIVO
0014-09-1978-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03
0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03	0003-09-0004-00-03

Figura 21: Imagen que muestra la base de datos primaria de la UT
Fuente: Base de datos primaria de la Unidad de Titulación de Ingeniería de Sistemas

En cuanto a los datos del periodo académico 55 deben ser obtenidos de las resoluciones pertinentes al estado del proceso. Mediante este proceso se realizaron las pruebas de funcionalidad del sistema y se pudo observar los problemas relacionados a este proceso.

3.2.1. Proceso de ingreso de temas, alumnos y lectores asignados en la Unidad de Titulación

El Coordinador de la Unidad de Titulación debe ingresar datos específicos en el sistema. Estos datos son de suma importancia para el proceso puesto que, a pesar de que no son ingresados mediante resoluciones, son relevantes para la continuidad del proceso de titulación.

Las siguientes son las tareas del Coordinador de la Unidad de Titulación:

- i. El Coordinador de la Unidad de Titulación tiene, dentro de la solución, un módulo de ingreso de documentos, en este espacio deberá ingresar los siguientes documentos pertinentes al Proceso de Titulación:
 - a. Listado de temas para proyectos de titulación (Proyecto Técnico o Artículo Académico) con sus respectivos tutores (docentes que proponen los temas).
 - b. Listado de Docentes asignados como lectores (lectores de plan o de borrador del proyecto de titulación)
 - c. Listado de alumnos que se inscriben a la Unidad de Titulación con los datos especificados. (Anexo 6)
 - d. Ingreso de resoluciones (contingencia de resoluciones).
- ii. El Coordinador de la Unidad de Titulación debe ingresar el listado de temas para proyectos de titulación. Esta lista debe estar en formato Excel y debe contener los datos relevantes del tema del proyecto de titulación, así mismo debe informar si estos temas están o no aprobados para su publicación.

- iii. El Coordinador de la Unidad de Titulación debe ingresar el listado de docentes que han sido determinados como lectores, tantos lectores de plan cómo lectores del borrador del proyecto de titulación.
- iv. El Coordinador de la Unidad de Titulación debe ingresar el listado de alumnos que se inscriben al Proceso de Titulación. Este listado está en formato Excel y contiene los datos relevantes del alumno e informa si la solicitud de ingreso a la Unidad de Titulación ha sido aprobada o negada.
- v. El Coordinador de la Unidad de Titulación tiene la potestad de ingresar ciertas resoluciones, en caso de que los alumnos no lo hayan realizado, esto con el fin de evitar contradicciones y vacíos en la data de la Unidad de Titulación y la solución de BI.

3.2.2. Proceso de ingreso de datos históricos

Una vez construido el data mart de la Unidad de Titulación, se procedió con el ingreso de datos históricos con el objetivo de realizar un análisis preliminar y la verificación del correcto funcionamiento del gestor de base de datos y la herramienta de visualización BI.

- i. En primer lugar, se realizó una revisión de la base de datos primaria, que manejaba la unidad de titulación descrita en la Figura 22. Esta base de datos se encontraba en formato .xls (libro de Excel) y fue otorgada por el Coordinador de la Unidad de Titulación de la carrera de ingeniería de Sistemas y Computación.

Base de datos primaria de la UT

OPCIÓN DE TITULACIÓN			CARRERA	REVISIÓN	CARRERA ANTIGÜEDAD	CARRERA	REVISIÓN	CARRERA ANTIGÜEDAD	CARRERAS				REVISIÓN	CARRERA ANTIGÜEDAD	REVISIÓN	REVISIÓN	REVISIÓN
1.	OPCIÓN DE TITULACIÓN	CARRERA	REVISIÓN	CARRERA ANTIGÜEDAD	CARRERA	REVISIÓN	CARRERA ANTIGÜEDAD	REVISIÓN	REVISIÓN	REVISIÓN	REVISIÓN	REVISIÓN	REVISIÓN	REVISIÓN	REVISIÓN	REVISIÓN	REVISIÓN
1.	EXAMEN COMPLEJO I	EXAMEN COMPLEJO I							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
2.	EXAMEN COMPLEJO II	EXAMEN COMPLEJO II							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
3.	EXAMEN COMPLEJO III	EXAMEN COMPLEJO III							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
4.	EXAMEN COMPLEJO IV	EXAMEN COMPLEJO IV							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
5.	EXAMEN COMPLEJO V	EXAMEN COMPLEJO V							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
6.	EXAMEN COMPLEJO VI	EXAMEN COMPLEJO VI							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
7.	EXAMEN COMPLEJO VII	EXAMEN COMPLEJO VII							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
8.	EXAMEN COMPLEJO VIII	EXAMEN COMPLEJO VIII							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
9.	EXAMEN COMPLEJO IX	EXAMEN COMPLEJO IX							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					
10.	EXAMEN COMPLEJO X	EXAMEN COMPLEJO X							2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00	2023-08-20 00:00					

Figura 22: Estructura de la base de datos primaria de la UT

Fuente: Base de datos primaria de la Unidad de Titulación de Ingeniería de Sistemas

- ii. Los datos que reposan en el archivo Excel fueron clasificados según los criterios obtenidos durante el análisis y modelamiento del data mart, de manera que se puedan clasificar a los involucrados y los parámetros dentro del proceso de titulación. Una vez realizada la clasificación se procedió con la limpieza y transformación, de manera manual, de los datos evitando la existencia de valores nulos o en blanco dentro de la solución, esto se ve reflejado en la Figura 23.

Base de datos primaria procesada.

ID_RESOLUCION	ID_TIPO_RESOLUCION	APROBADO	FECHA_RESOLUCION	
0	0	1	1/1/1900	
0	0	0	1/1/1900	
0	0	0	1/1/1900	
0	0	0	1/1/1900	
00228-009-2018-04-09		1	1/1/1900	
23151-027-2017-11-06		2	1/1/1900	
18463-006-2015-03-30		5	1	30/3/2015
18414-006-2015-03-30		5	1	30/3/2015
18446-006-2015-03-30		5	1	30/3/2015
18435-006-2015-03-30		5	1	30/3/2015
18416-006-2015-03-30		5	1	30/3/2015
18432-006-2015-03-30		5	1	30/3/2015
18413-006-2015-03-30		5	1	30/3/2015
18410-006-2015-03-30		5	1	30/3/2015
18411-006-2015-03-30		5	1	30/3/2015
18451-006-2015-03-30		5	1	30/3/2015
18407-006-2015-03-30		5	1	30/3/2015
18475-006-2015-03-30		5	1	30/3/2015
18428-006-2015-03-30		5	1	30/3/2015
18448-006-2015-03-30		5	1	30/3/2015
18417-006-2015-03-30		5	1	30/3/2015
18444-006-2015-03-30		5	1	30/3/2015
18419-006-2015-03-30		5	1	30/3/2015
18457-006-2015-03-30		5	1	30/3/2015
18465-006-2015-03-30		5	1	30/3/2015
18464-006-2015-03-30		5	1	30/3/2015
18443-006-2015-03-30		5	1	30/3/2015

► ... | D_ALUMNO | D_TEMA | **D_RESOLUCION** | D_DOCENTE | D_OPCION | D_PERIODO | D_INFORME | TIPO_INFORME | ... ⊕ :

Figura 23: Imagen que muestra la base de datos primaria de la UT una vez que fue procesada (estandarizada) manualmente.

Elaborado por: Jhonny Bermeo y Edison Simbaña.

- iii. Los datos están listos para ser ingresados en la BDD relacional, posterior a esto serán cargados desde el gestor de BDD MySQL como se muestra en la Figura 24.

Comandos SQL necesarios para el ingreso de datos en la BDD

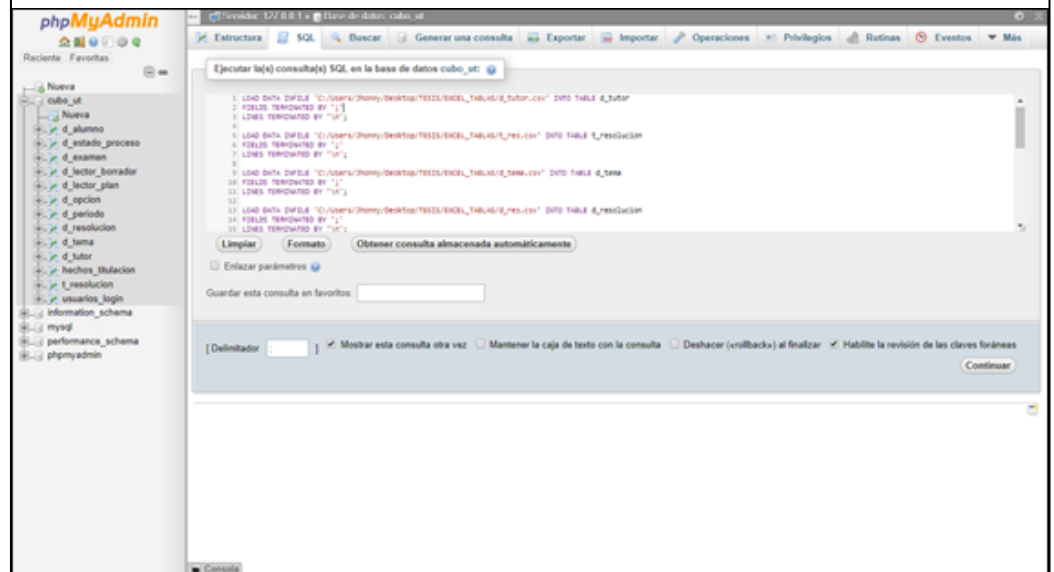


Figura 24: Imagen que muestra los comandos SQL necesarios para el ingreso de datos en la BDD almacenada en el gestor MySQL

Elaborado por: Jhonny Bermeo y Edison Simbaña.

- iv. Una vez cargados los datos históricos limpios en la BDD, se procede con la creación de la conexión entre este gestor y la herramienta de análisis descrita en la Figura 25, esto se establece mediante la herramienta Data Source Management que Pentaho provee en su versión de la Comunidad (Community Edition).

Conexión Pentaho-MySQL

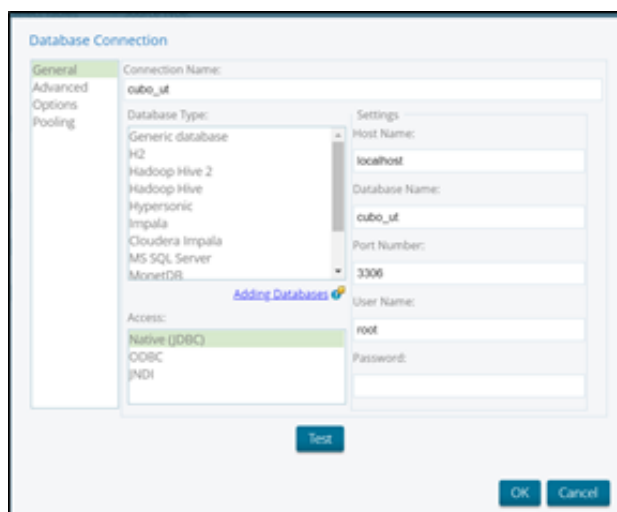


Figura 25: Imagen que muestra la creación de la conexión Pentaho-MySQL con la herramienta Data Source Wizard

Fuente: Software pentaho.

- v. Una vez establecida la conexión con la base de datos, se procede con la creación del cubo OLAP con la herramienta Schema Workbench la cual es provista por la suite de Pentaho en su versión de la Comunidad descrita en la Figura 26. En esta herramienta se crean las dimensiones, la tabla de hechos y las medidas en base al análisis realizado y finalmente se procede a publicar el cubo OLAP realizado y del cual Pentaho se va a valer para realizar el análisis visual.

Herramienta Schema Workbench.

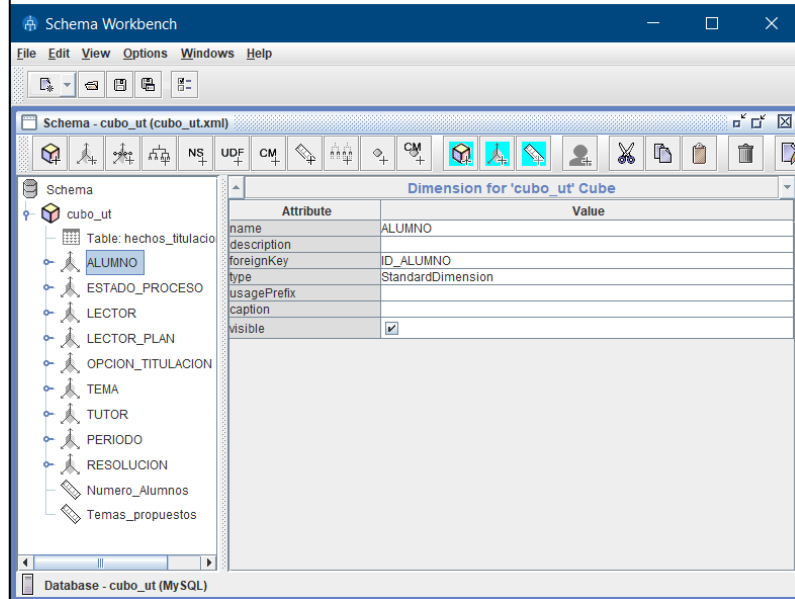


Figura 26: Imagen que muestra la interfaz de la herramienta Schema Workbench.

Fuente: herramienta Schema Workbench.

- vi. Con ayuda de la herramienta Saiku Analytics, provista por la suite de Pentaho en su versión de la Comunidad en la Figura 27, se procede a realizar las consultas de análisis sobre los datos históricos ingresados en la base de datos relacional.

- e. Tema del proyecto de titulación (Para estudiantes inscritos en la opción Examen Complexivo (E) se atribuye el valor como “NO APLICA”)
 - f. Nombre del tutor/lector (Para estudiantes inscritos en la opción Examen Complexivo (E) se atribuye el valor como “NO APLICA”)
- ii. Una vez extraídos los datos necesarios de la resolución se procede con la validación y verificación:
- a. Validar si el número de cedula del alumno está ingresado en la tabla dimensional de los alumnos, lo que indica que ya se encuentra inscrito en la unidad de titulación y, por ende, está presente en el data mart.
 - b. Verificar que el tema del proyecto esté ingresado en la tabla dimensional de los temas, lo que indica que el tema fue aprobado y ya se encuentra en el data mart, se procede con la obtención de su ID. (Esto aplica únicamente para los casos de proyecto técnico o artículo académico)
 - c. Verificar que el nombre del tutor/lector se encuentre dentro del data mart y obtener su ID (Esto aplica únicamente para los casos de proyecto técnico o artículo académico)
- iii. Luego de la validación de datos se procede con la transformación de estos:
- a. Se verifica que no existan caracteres especiales en los datos.
 - b. Cambiar los caracteres especiales por caracteres coherentes para que el dato no pierda valor
 - c. Estandarizar los datos llevando los caracteres alfabéticos a mayúsculas.
- iv. Una vez se han transformado los datos se procede con su carga en el data mart, tomando en cuenta casos especiales, se debe realizar de la siguiente manera:

- a. El número, el tipo y la fecha de la resolución son ingresadas a su respectiva dimensión.
- b. Se ingresa en la tabla de hechos los campos ID, obtenidos de las dimensiones durante el proceso ii.b y ii.c, según corresponda y se atribuye el estado del proceso según corresponda.
- c. Casos especiales

A lo largo del proceso ETL llevado a cabo en las resoluciones disponibles del periodo 55, se encontraron varios casos especiales que se deben considerar para tener un óptimo desempeño de la solución.

3.2.3.1. Estudiantes que ingresan al curso de actualización

Los estudiantes que desean ingresar a la Unidad de Titulación tras 1 año de su egreso, o tras 1 año de no haber culminado con su proyecto de titulación deben primero concluir con el curso de actualización, esto indica que los estudiantes pueden o no estar dentro de la dimensión de estudiantes y, por ende, estar o no dentro del data mart, lo que conlleva a un caso especial de este tipo.

Para la resolución de este caso especial se lo realizó una diferenciación de aquellos estudiantes que están en esta situación y de los que no utilizando dos módulos web distintos para cada caso. Se debe acotar que los estudiantes, estén o no dentro de este caso especial, ingresan a sus respectivos módulos desde un único módulo de acceso (login).

3.2.3.2. Resoluciones faltantes

Este caso en específico conlleva un gran problema para la solución de inteligencia de negocios. El hecho de que un alumno no ingrese una resolución a tiempo conlleva a que los datos que esta resolución proveen, para el posterior análisis de datos, se

encuentren vacíos y formen una brecha de información la cual puede llevar a un incorrecto análisis de datos.

La solución para este tipo de problemas llegó en forma de un plan de contingencia para casos de este estilo.

3.2.3.3. Plan de contingencia en caso de resoluciones faltantes

Las resoluciones que no hayan sido subidas por el alumno en el estado del proceso correspondiente pueden ser cargadas a la solución por medio del Coordinador de la Unidad de Titulación quien tiene la potestad de subir este tipo de documentos en el módulo pertinente.

- i. Se considera que una resolución no fue subida, cuando el estado del proceso en curso llega a su fin y hay una, o varias resoluciones, correspondientes a ese estado del proceso, que no han sido subidas al repositorio.
- ii. La brecha provista por la falta de datos puede provocar una reportería deficiente de datos y conlleva a que el análisis de estos sea erróneo.

Una vez que el Coordinador de la Unidad de Titulación obtenga las resoluciones faltantes (en secretaria del campus) debe ingresarlas al sistema mediante la Aplicación Web que conforma el módulo ETL.

3.3. Explotación de información en una solución de BI

La explotación de información de la solución de BI de la Unidad de Titulación se puede evidenciar en los reportes, los cuales se obtienen con ayuda de la herramienta Saiku Analytics, herramienta propia de la suite de Pentaho CE, determinando así dos tipos de reportes:

a. Reportes fijos

Los reportes fijos son reportes que han sido determinados según la necesidad de análisis de la Unidad de Titulación. Dichos reportes se almacenan en el Coordinador de la solución y pueden ser accedidos por todos aquellos que conforman la Unidad de Titulación (Coordinador y docentes de apoyo).

El número de reportes fijos fue establecido en 3 los cuales son:

- **Historial del alumno**

Este reporte representado en la Figura 28 muestra, de manera tabulada, los eventos relevantes del alumnado con respecto al proceso de titulación. Consta de 5 columnas en las cuales se puede observar el nombre del alumno, el número y el tipo de resolución, el estado del proceso y el periodo académico.

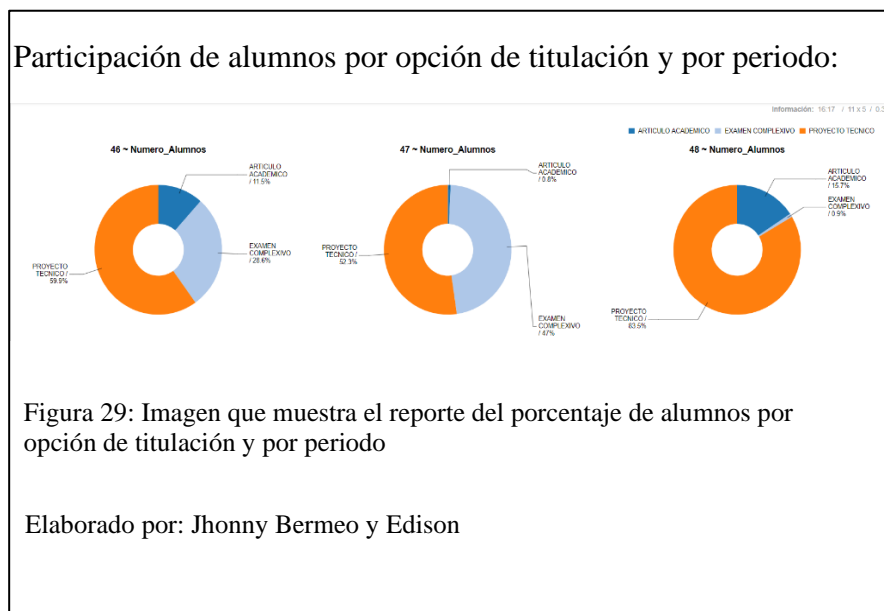
Historial del alumno.				
ALUMNO	NUMERO_RESOLUCION	TIPO_RESOLUCION	ESTADO_PROCESO	PERIODO
ABATA QUINCHUQUI CHRISTIAN PAUL	00235-009-2018-04-09	RESOLUCION DE 1RA PRORROGA	GESTION/PREPARACION	52
	00398-015-2018-06-18	RESOLUCION DE ENTREGA DE 1ER BORRADOR	GESTION/PREPARACION	52
	00504-020-2018-08-06	NO INFORMACION DISPONIBLE	FINALIZACION	52
ABRIL CASTILLO ROBERTO FERNANDO	23030-024-2017-09-25	RESOLUCION DE INGRESO A LA UT	INICIO_PROCESO	51
	18828-018-2015-09-02	RESOLUCION DE INGRESO A LA UT	INICIO_PROCESO	46
	999	NO INFORMACION DISPONIBLE	FINALIZACION	47
ACEVEDO ACEVEDO SANTIAGO DAVID	18407-006-2015-03-30	RESOLUCION DE INGRESO A LA UT	INICIO_PROCESO	46
	19004-020-2015-09-28	RESOLUCION DE 1RA PRORROGA	GESTION/PREPARACION	48
	19561-008-2016-04-11	RESOLUCION DE 2DA PRORROGA	GESTION/PREPARACION	48
ACHIG AGUILAR ROGER FERNANDO	22599-011-2017-04-10	RESOLUCION DE INGRESO AL CURSO DE ACTUALIZACION	CURSO ACTUALIZACION	48
	00086-004-2018-02-19	NO INFORMACION DISPONIBLE	FINALIZACION	51
	22547-010-2017-03-31	RESOLUCION DE INGRESO A LA UT	INICIO_PROCESO	50
ACHIG GUACHAMIN KAREN VANESSA	23062-024-2017-09-25	RESOLUCION DE 1RA PRORROGA	GESTION/PREPARACION	51
	23172-028-2017-11-20	RESOLUCION DE ENTREGA DE 1ER BORRADOR	GESTION/PREPARACION	51
	00033-002-2018-01-22	RESOLUCION DE ENTREGA DE 1ER BORRADOR	GESTION/PREPARACION	51
ACONDA ANALUIZA CARLOS RUBEN	00146-006-2018-03-08	NO INFORMACION DISPONIBLE	FINALIZACION	51
	22540-010-2017-03-31	RESOLUCION DE INGRESO A LA UT	INICIO_PROCESO	50
	23060-024-2017-09-25	RESOLUCION DE 1RA PRORROGA	GESTION/PREPARACION	51
ACOSTA GALINDO FREDDY ALFREDO	00245-009-2018-04-09	RESOLUCION DE 1RA PRORROGA	GESTION/PREPARACION	53
	00656-024-2018-10-09	RESOLUCION DE 2DA PRORROGA	GESTION/PREPARACION	53
	23034-024-2017-09-25	RESOLUCION DE INGRESO A LA UT	INICIO_PROCESO	51
ACOSTA GALINDO FREDDY ALFREDO	18406-006-2015-03-30	RESOLUCION DE INGRESO A LA UT	INICIO_PROCESO	46
	19486-007-2016-03-29	RESOLUCION DE 1RA PRORROGA	GESTION/PREPARACION	48
	19535-008-2016-04-11	RESOLUCION DE 2DA PRORROGA	GESTION/PREPARACION	48
	22123-023-2016-10-24	RESOLUCION DE INGRESO AL CURSO DE ACTUALIZACION	CURSO ACTUALIZACION	48
	22346-003-2017-01-30	RESOLUCION DE ENTREGA DE 1ER BORRADOR	GESTION/PREPARACION	48
22470-009-2017-03-27	RESOLUCION FINAL DE APTITUD	FINALIZACION	48	

Figura 28: Imagen que muestra el reporte fijo del historial de los alumnos.

Elaborado por: Jhonny Bermeo y Edison

- **Participación de alumnos por opción de titulación y por periodo.**

Este reporte fijo representado en la Figura 29 muestra, de manera gráfica, el porcentaje de participación de los alumnos por cada opción de titulación desde el periodo No. 46



- **Temas propuestos por tutor y por opción en cada periodo**

De manera tabulada, el reporte de la Figura 30 muestra la cantidad de temas que un docente propuso durante un determinado periodo, diferenciando si estos temas fueron para artículos académicos o para proyectos técnicos.

Temas propuestos por periodo y por tipo de proyecto de titulación

TUTOR	PERIODO	OPCION_TITULACION	Temas_propuestos
NO APLICA	46	EXAMEN COMPLEXIVO	1
	47	EXAMEN COMPLEXIVO	1
	48	EXAMEN COMPLEXIVO	1
	49	EXAMEN COMPLEXIVO	1
	51	EXAMEN COMPLEXIVO	1
AGUAYO MORALES JOSE LUIS	46	PROYECTO TECNICO	2
	47	PROYECTO TECNICO	4
	48	PROYECTO TECNICO	6
	49	PROYECTO TECNICO	2
	51	PROYECTO TECNICO	1
AREVALO CAMPOS ALONSO RENE	53	PROYECTO TECNICO	1
	46	PROYECTO TECNICO	6
	47	PROYECTO TECNICO	4
	48	PROYECTO TECNICO	3
	49	PROYECTO TECNICO	3
DIAZ ORTIZ DANIEL GIOVANNY	50	PROYECTO TECNICO	6
	51	PROYECTO TECNICO	2
	52	PROYECTO TECNICO	1
	53	PROYECTO TECNICO	2
	999	PROYECTO TECNICO	1
	46	PROYECTO TECNICO	4
	47	PROYECTO TECNICO	4
	48	PROYECTO TECNICO	4
	49	PROYECTO TECNICO	5
	50	PROYECTO TECNICO	8
	51	PROYECTO TECNICO	6
			5
			2

Figura 30: Imagen que muestra el reporte del número de temas propuestos por periodo y por tipo de proyecto de titulación

Elaborado por: Jhonny Bermeo y Edison

a. Reportes bajo demanda

Con ayuda de la herramienta Saiku Analytics en la Figura 31, se pueden crear reportes dinámicos bajo demanda. Aquellas personas que conforman la Unidad de Titulación pueden crear y visualizar este tipo de reportes bajo demanda, además de poder almacenarlos.

Herramienta de análisis Saiku Analytics



CUTTING EDGE OPEN SOURCE ANALYTICS

Saiku has the power to change the way you think about your business and make decisions. Saiku provides powerful, web based analytics for everyone in your organisation. Quickly and easily analyse data from any data source to discover what is really happening inside and outside your organisation.

Quick Links



Figura 31: Imagen que muestra la página de inicio de la herramienta Saiku Analytics de Sistemas Campus Sur

Elaborado por: Jhonny Bermeo y Edison

Para la creación de reportes con la herramienta Saiku, se debe acceder a la opción **CREATE A NEW QUERY** para luego proceder a crear la consulta respectiva con las herramientas proporcionadas.



Figura 32: Panel de creación de reportes de la herramienta Saiku Analytics

Fuente: Herramienta Saiku Analytics

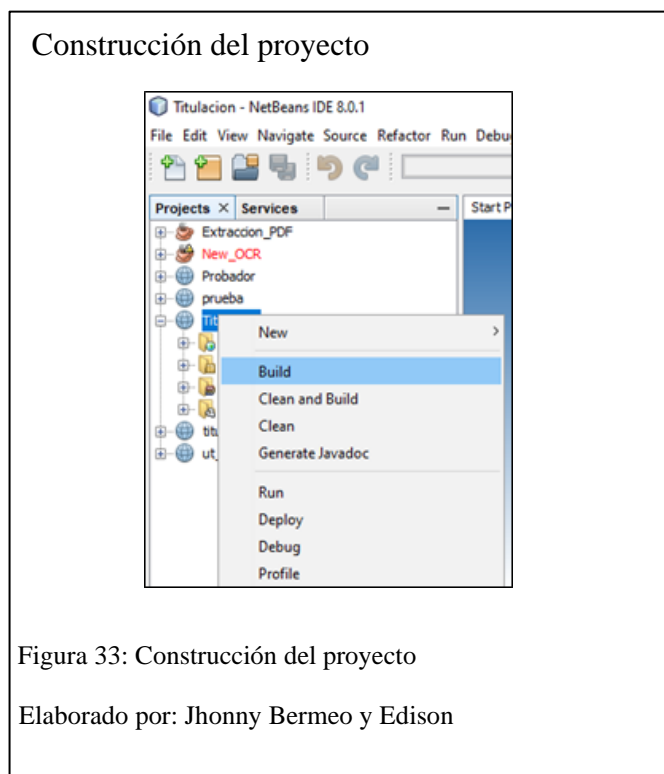
Nota: Para la creación de un reporte en el panel mostrado en la Figura 32 se debe verificar que haya mínimo una columna y una fila o, en su defecto, mínimo una medida y una fila o columna, para que la herramienta muestre la información requerida.

3.4. Levantamiento de la solución en el servidor

Una vez que la solución fue probada se procedió al levantamiento de esta en el servidor designado:

Componente Web ETL

1. Crear el archivo .war que será cargado en el servidor de aplicaciones Tomcat. Para esto se debe construir (Build) el proyecto en este caso en el IDE Netbeans, seleccionando el proyecto y dando clic izquierdo para seleccionar la opción como se muestra en la Figura 33.



2. El archivo .war después de realizar la construcción del proyecto se alojará en la carpeta **dist** en este caso Titulacion.war como se muestra en la Figura 34.

Archivo .war

build	13/01/2020 19:22	Carpeta de archivos
dist	13/01/2020 19:22	Carpeta de archivos
Lib_extraccion	13/01/2020 18:26	Carpeta de archivos
nbproject	13/01/2020 19:11	Carpeta de archivos
src	13/01/2020 18:26	Carpeta de archivos
web	13/01/2020 18:26	Carpeta de archivos
build	13/01/2020 19:13	Documento XML

Figura 34: Directorio de localización de archivo .war

Elaborado por: Jhonny Bermeo y Edison

3. El archivo .war es desplegado en el servidor Tomcat mediante el módulo de administración de aplicaciones representado en la Figura 35, para acceder al mismo se debe configurar el archivo tomcat-users.xml y adicionar el usuario y contraseña con el rol **manager-gui**.

Gestor de aplicaciones web

Aplicación	Versión	Modulo a Monitorear	Estado	Sesiones	Comandos
/	Ninguno especificado	Welcome to Tomcat	Yes	0	Avanzar Parar Reiniciar Detener [Explora sesiones] [Ver trabajos] [0] [Iniciar]
chibacco	Ninguno especificado		Yes	0	Avanzar Parar Reiniciar Detener [Explora sesiones] [Ver trabajos] [0] [Iniciar]
chibacco2	Ninguno especificado		Yes	0	Avanzar Parar Reiniciar Detener [Explora sesiones] [Ver trabajos] [0] [Iniciar]
docs	Ninguno especificado	Tomcat Documentation	Yes	0	Avanzar Parar Reiniciar Detener [Explora sesiones] [Ver trabajos] [0] [Iniciar]
examples	Ninguno especificado	Servlet and JSP Examples	Yes	0	Avanzar Parar Reiniciar Detener [Explora sesiones] [Ver trabajos] [0] [Iniciar]
examples2	Ninguno especificado	Tomcat Host Manager Application	Yes	0	Avanzar Parar Reiniciar Detener [Explora sesiones] [Ver trabajos] [0] [Iniciar]
manager	Ninguno especificado	Tomcat Manager Application	Yes	0	Avanzar Parar Reiniciar Detener [Explora sesiones] [Ver trabajos] [0] [Iniciar]

Figura 35: Gestor de aplicaciones web del servidor Tomcat

Elaborado por: Jhonny Bermeo y Edison

4. En caso de que el archivo .war tenga un tamaño mayor a 50Mb es necesario configurar el archivo server.xml y aumentar el tamaño del parámetro maxPostSize como se establece en la Figura 36.

Archivo server.xml

```
<Connector port="8080" protocol="HTTP/1.1"
  connectionTimeout="20000"
  maxPostSize="104857600"
  redirectPort="8443" />
<!-- A "Connector" using the shared thread pool-->
<!--
```

Figura 36: Configuración archivo server.xml de tomcat

Elaborado por: Jhonny Bermeo y Edison

La aplicación web ya puede ser accedida mediante la IP:

172.17.42.177:8080/Titulacion.

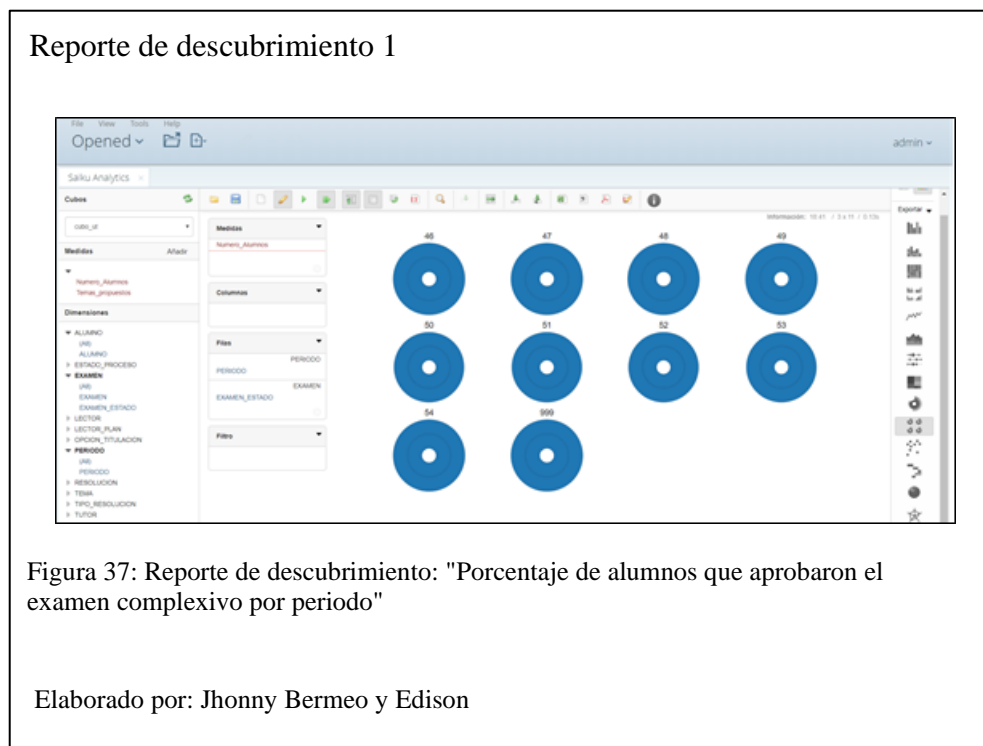
3.5. Descubrimientos

En base a la reunión mantenida con un miembro de la Unidad de Titulación, quien tiene conocimiento del proceso de titulación, se realizaron los siguientes descubrimientos que aumentan el catálogo de reportes de BI en el servidor.

Reportes solicitados que se realizaron sin inconvenientes:

1. Porcentaje de alumnos que aprobaron el examen complejo por periodo.

Este descubrimiento involucra las dimensiones examen, periodo y a la medida Alumnos. Permite observar, gráficamente, el porcentaje de alumnos que aprobaron el examen complejo segregados por periodo representados en la Figura 37.



2. Cantidad de veces que un profesor fue elegido lector en varios periodos.

Este descubrimiento involucra a las dimensiones lector, periodo y a la medida Temas representados en la Figura 38. Mediante este reporte se puede observar la cantidad de proyectos de titulación que un docente ha leído en un determinado periodo.

Reporte de descubrimiento 2

The screenshot shows a software interface for data analysis. On the left, there is a 'Dimensiones' (Dimensions) pane with a tree view containing categories like ALUMNO, ESTADO_PROCESO, EXAMEN, LECTOR, LECTOR_PLAN, LECTOR_PLAN_PLAN, LECTOR_PLAN_PLAN_PLAN, LECTOR_TITULACION, PERIODO, PERIODO_PLAN, RESOLUCION, TEMA, TPOD_RESOLUCION, and TUTOR. The 'Medidas' (Measures) pane on the right contains 'Temas'. The main area displays a pivot table with 'LECTOR' and 'PERIODO' as row labels and 'Temas' as column labels. The table shows counts for various teachers across different periods.

LECTOR	PERIODO	Temas
NO APLICA	46	26
	47	10
	48	10
	49	4
	50	4
	51	11
	52	6
	53	10
	54	3
	55	5
AGUIRRE MORALES JOSE LUIS	47	3
	48	2
	49	3
	50	2
	51	2
	52	2
	53	1
	54	1
	55	1
AREVALO GARCIA ALONSO RENE	46	4
	47	1
	48	4
	49	4
	50	4
	51	4
	52	1
	53	1
	54	3
	55	3

Figura 38: Reporte de descubrimiento: "Cantidad de veces que un profesor fue elegido lector en varios periodos"

Elaborado por: Jhonny Bermeo y Edison

3. Estudiantes que han ido a Curso de Actualización en determinado periodo.

Con este descubrimiento se puede observar el número de alumnos en curso de actualización por periodo académico. Involucra las dimensiones Alumno, Estado del Proceso y Periodo representados en la Figura 39, el reporte permite observar, además, si un alumno asistió a más de un curso de actualización.

Reporte de descubrimiento 3

The screenshot shows a data analytics application interface. On the left, there is a sidebar with a tree view of dimensions and measures. The main area displays a data table with the following columns: ESTADO_PROCESO, ALUMNO, PERIODO, and alumno. The table contains 30 rows of data representing students and their course update status.

ESTADO_PROCESO	ALUMNO	PERIODO	alumno
CURSO_ACTUALIZACION	ACOSTA ACYREDO SANTIAGO DAVID	48	1
	ACOSTA SALAZAR FREDY ALFONSO	48	1
	ADULA FLORES CARLOS OSWALDO	399	1
	ARANA SANCHEZ RICARDO MARCELO	48	1
	BOLANDES ALDAS ADOYANA MARGARITA	48	1
	BROWN LAICA RICHARD FABRIZIO	48	1
	CHURRO HERNANDEZ WILLIAM GONZALO	48	1
	CRIQUE GONZALEZ SHARINA ALEXANDER	48	1
	DAWILA DARLA JAIME	48	1
	DE LA TORRE GORDON NELSON DAVID	48	1
	ESPINOZA HERNANDEZ OSWALDO ALEJANDRO	47	1
	FLORES GONZALEZ CHRISTIAN FABIAN	50	1
	FLORES HERNANDEZ MONICA ALEXANDRA	48	1
	FRANCO HERNANDEZ JUAN CARLOS	50	1
	GARCIA BARRAGAN MARIA DEL CARMEN	50	1
	SUMARIBINO TARECO DIEGO SAM	54	1
	SUMARIBINO TOSCAN CAROLINA ELIZABETH	48	1
	HERRERA BALBUENA CARLOS ANDRES	50	1
	MORETA TRUJILLO WELLINGTON GUTEMBERG	48	1
	MUNEZ ZAMBRANO JORGE FERNANDO	399	1
	MURRAY SENA GARCIA FRANCISCO GABRIEL	48	1
	PARVEDO CALDERON RAULITO RENE	48	1
	PARVEDO SAMARINO RODRIGO ALFONSO	399	1
	QUILLO JUAN PABLO ELIZABETH ROJO	48	1
	RODRIGUEZ CALVOFRAN VERONICA ELIZABETH	48	1
	RODRIGUEZ CORTES EDGAR LEONARDO	48	1
	RODRIGUEZ LARCO MONICA VIVIANA	50	1
	SAHADO OLIVERA ANDREA ALEXANDER	50	1

Figura 39: Reporte de descubrimiento: "Estudiantes que han ido a Curso de Actualización en determinado periodo."

Elaborado por: Jhonny Bermeo y Edison

4. Estudiantes que han solicitado 2da prorroga en determinado periodo.

No se puede saber con certeza que alumno solicitó una 2da prorroga, debido a que la solución maneja únicamente resoluciones mas no solicitudes. Sin embargo, se puede realizar esta consulta tomando en cuenta a aquellos alumnos que tienen una resolución de 2da prorroga aprobada y subida al data mart como se muestra en la Figura 40. Esta consulta involucra las dimensiones alumno, periodo y resolución sobre la cual se aplica el filtro correspondiente.

Reporte de descubrimiento 4

TIPO_RESOLUCION	ALUMNO	PERIODO	Alumnos
RESOLUCION DE 2da PRORROGA	ALFREDO ACEVEDO SANTIAGO DAVID	48	1
	ACENDA MALLERA CARLOS RUBEN	53	1
	ACOSTA GALINDO FREDDY ALFREDO	48	1
	ALBA SOLERINO JORGE LUIS	53	1
	ALBIA TAPIA SORIA DE LOS ANGELES	48	1
	ALMESA MUJER JONATHAN FERNANDO	53	1
	ALVARADO CORTES ALEJANDRO EBRAIM	53	1
	ALVAREZ ANDRÉS EDUARDO RAMIRO	48	1
	ANGAMARCA PUNALES OSCAR WASHINGTON	48	1
	ANWARO CHEN LUIS MIGUEL	53	1
	ARANA HERRERA JONATHAN ELIASER	48	1
	ARAUJO ESCOBAR EDUARDO ALFONSO	53	1
	ARROYO CADA JONATHAN DANIEL	53	1
	ARAYA SANCHEZ RICARDO SAMPURNO	48	1
	ARIBE TOBIAS JUAN CARLOS	53	1
	BALBEDA RODRIGUEZ CARLOS ALEJANDRO	48	1
	BONILLA LARCO FRANCISCO JAVIER	48	1
	BRAND POLCHEROS CHRISTIAN DANIEL	53	1
	BRANG SANCHEZ TAMICO VINICIO	53	1
	CAJAMARCA BENITEZ BYRON DANIEL	48	1
	CALZADILLA BARRERA EDWIN FERNANDO	47	1
	CABRILLO VILLEGAS DIEGO JAVIER	54	1
	CABA LOPEZ LUIS FERNANDO	48	1
	CELI RAMIREZ CATALINA MARIBEL	53	1
	CHANGIZ SAAVEDRA DAFIN YANISRA	48	1
	CHANGIZ CHANGIZ LARA DIANA CAROLINA	53	1
	CHAVEZ SANCHEZ CARLOS ALBERTO	53	1
	CHARRINOLA VENTURINI JAVIER DANIEL	48	1

Figura 40: Reporte de descubrimiento: "Estudiantes que han solicitado 2da prorroga en determinado periodo"

Elaborado por: Jhonny Bermeo y Edison

Reportes solicitados donde se encontraron inconvenientes:

Por otro lado, se tiene el reporte que, debido a cuestiones del modelo del data mart, no se pudo realizar en un principio, sin embargo, al optimizar el modelo dimensional del data mart se ha podido realizar.

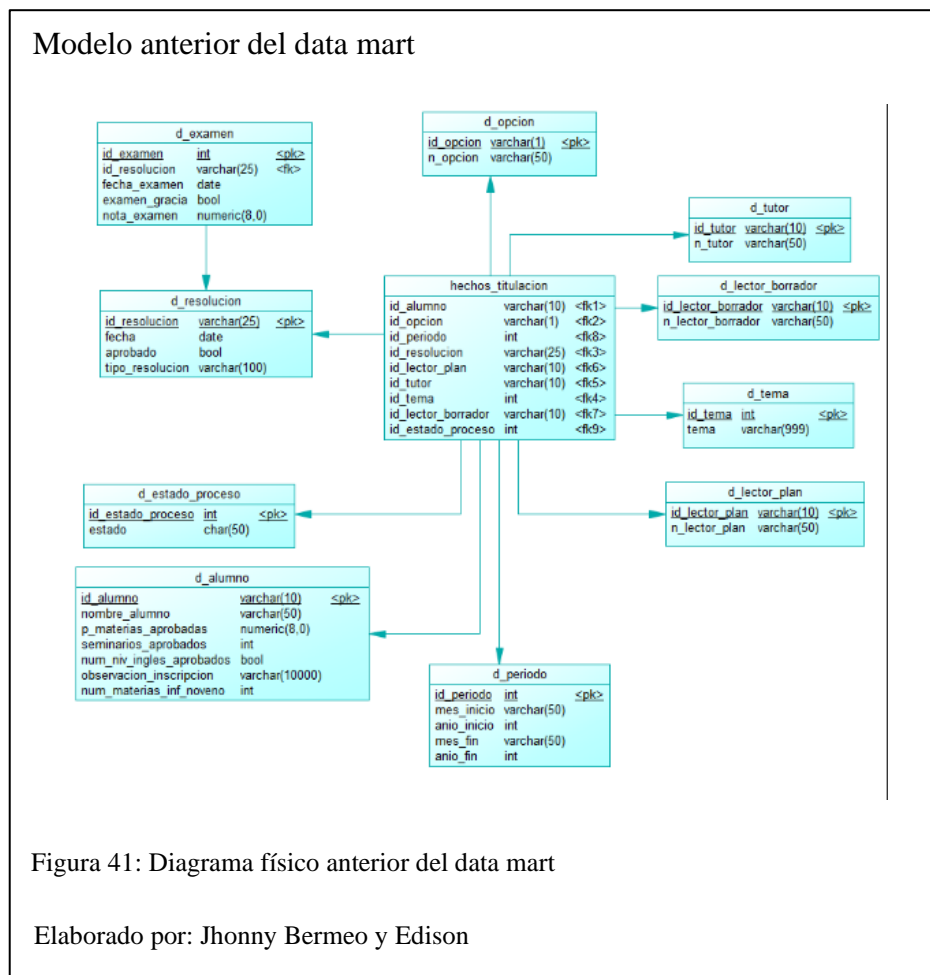
Para realizar este mejoramiento al data mart, en primer lugar, se descubrió que es lo que estaba impidiendo obtener el reporte deseado.

Debido a que Pentaho y su herramienta de reportes Saiku Analytics trabajan con el cubo OLAP y realizan consultas que involucran directamente a la tabla de hechos y a las tablas dimensionales, se encontró la limitación en el modelo del data mart que se muestra en la Figura 41.

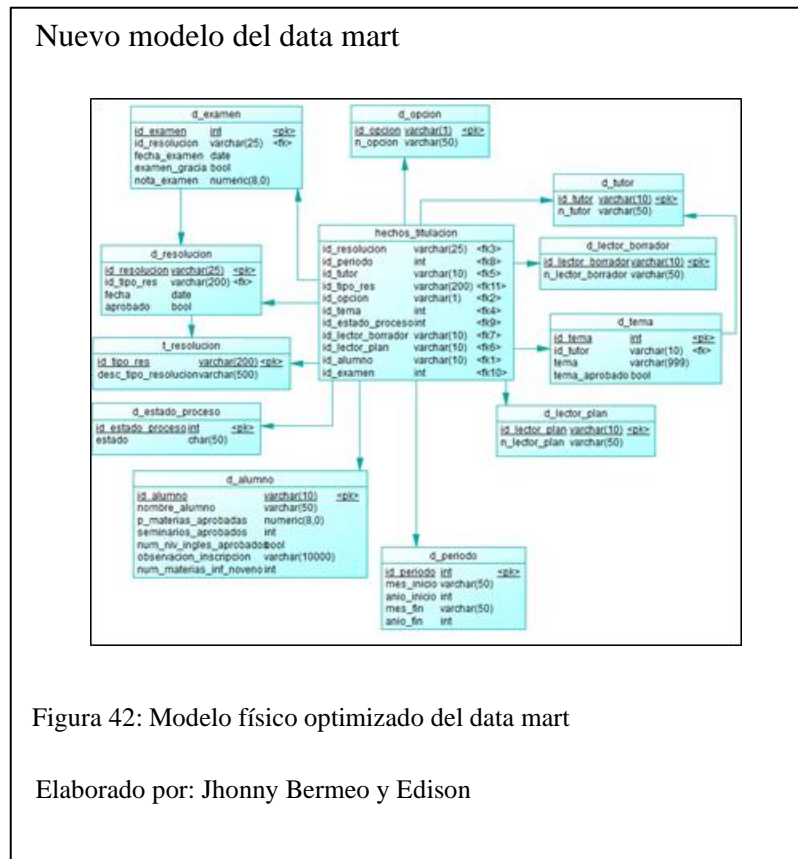
Una vez identificado el problema se procedió con el mejoramiento que consiste en agregar una nueva columna a la tabla de hechos y relacionarla con la columna del

identificador primario de la tabla “tipo de resolución”, con esto se crea una dimensión adicional al modelo que cumple dos funciones: la primera es ser la tabla dimensión “Tipo de resolución” y la segunda es proveer el tipo de resolución a la dimensión “Resolución”.

Este último paso está totalmente permitido debido a que el modelo dimensional elegido es el copo de nieve que permite la relación entre tablas dimensionales, tablas no dimensionales y la tabla de hechos.



El modelo físico del data mart quedaría finalmente de la siguiente manera representado en la Figura 42:



Finalmente, se procedió con la creación del reporte en cuestión presentado en la Figura 43.

1. Estudiantes que han cambiado de tutor en su proceso de titulación.

Para realizar esta consulta se debe involucrar las dimensiones alumno, tipo de resolución y periodo, para posteriormente aplicar el filtro sobre la dimensión tipo de resolución y, si así se desea, aplicar el filtro sobre la dimensión periodo para obtener los resultados deseados.

Reporte de descubrimiento 5

The screenshot shows a Saku Analytics interface with the following components:

- Measures:** Alumnos
- Dimensions:** ALUMNO, ESTADO_PROCESO, EXAMEN, LECTOR, LECTOR_PLAN, OPCION_TITULACION, PERIODO, RESOLUCION, TEMA
- Table Data:**

TIPO_RESOLUCION	ALUMNO	PERIODO	Alumnos
CAMBIO DE TUTOR DE PROYECTOS DE TITULACION	DAVILA DAVILA JAJME	45	1
	MENA CARRASCO NELLY TATHIANA	45	1

Figura 43: Reporte de descubrimiento: " Estudiantes que han cambiado de tutor en su proceso de titulación"

Elaborado por: Jhonny Bermeo y Edison

CONCLUSIONES

- Se estableció que las fuentes primarias de datos para el proceso de titulación son las resoluciones puesto que estas marcan tanto el inicio como el fin de cada fase del proceso y contienen información relevante para el proceso de titulación.
- Los reportes fijos, creados con la herramienta Saiku Analytics, proporcionan información sobre el estado de los alumnos en el proceso de titulación y permiten observar un resumen porcentual e histórico sobre las opciones de proyecto de titulación en el transcurso de los últimos periodos académicos.
- Se desarrollo un módulo ETL que, tras la carga de archivos al servidor, permite la extracción de información de los documentos relevantes para el proceso de titulación, mediante técnicas de extracción de datos de texto y extracción de caracteres por medio de OCR. (Optical Character Recongnition)
- Se implementó el data mart de modelo copo de nieve debido a la naturaleza relacional de la base de datos de la Unidad de Titulación, que se obtuvo tras el análisis realizado a la base de datos primaria que manejaba dicha entidad.
- El servidor Glassfish para JSF no es compatible con soluciones web que conllevan estilos de presentación diferentes a los proporcionados por el framework Java Prime Faces además que presenta incompatibilidad con librerías OCR utilizadas en este proyecto, por lo cual el módulo ETL fue levantado en un servidor tipo Tomcat.
- La versión de la comunidad de Pentaho (Pentaho CE) es una herramienta que ayuda con la administración y visualización de reportes, sin embargo, no tiene herramientas propias integradas para la creación de estos, por ende, es

importante adquirir herramientas para estas tareas, como Saiku Analytics, en su Marketplace integrado.

- Los descubrimientos realizados permitieron el mejoramiento del modelo del data mart manteniendo la estructura copo de nieve que originalmente fue adaptada para la creación del mismo.

RECOMENDACIONES

- La revisión de los estándares planteados para los formatos de ingreso de información por parte de la Unidad de Titulación debe tener carácter riguroso dado que la omisión de estos puede acarrear inconcordancia de datos además de errores de inserción a la base de datos.
- Para el levantamiento de soluciones web con Java Server Faces, es recomendable utilizar servidores de tipo Tomcat, debido a que estos permiten la integración de todas las librerías externas, así como su correcta ejecución en ambiente web.
- Para trabajos futuros se recomienda el mejoramiento del servidor BI Pentaho CE por una versión profesional de pago, debido a que esta presentaría mejoras en la visualización y creación de los reportes.
- El proyecto plantea una escalabilidad hacia nuevos tipos de formatos de documentos y su lectura para manejar subprocesos manuales que podrían ser integrados en una futura automatización.

GLOSARIO.

BI: Bussinnes Intelligence o Inteligencia de Negocios. Es un proceso que consiste en obtener, recopilar, analizar y presentar información relevante para la toma de decisiones. (Kimball & Ross, 2013)

BPMN: Business Process Management Notation o Notación de Modelado de Procesos de Negocio. Es una forma estandarizada de representar procesos de negocio para propósitos descriptivos de alto nivel. (White & Miers, 2016)

ETL: Extract, Transform & Load o Extracción, Transformación y Carga. Es un proceso que consiste en una serie de procedimientos para obtener los datos de un repositorio, limpiarlos o estandarizarlos (transformarlos) y cargarlos en la base de datos de negocio. (Kimball & Ross, 2013)

Data mart: Es una versión más pequeña de los data warehouse, se los denomina también como data warehouse departamentales. A diferencia de los data warehouse, los data mart poseen un nivel de granularidad mayor puesto que solo se enfocan en un área, o proceso, de la organización. (Kimball & Ross, 2013)

Data warehouse: Data warehouse es una base de datos corporativa, desnormalizada no modelada en base a los sistemas transaccionales, que es alimentada por todos los departamentos que conforman la organización y usualmente están modelados en base a esquemas multidimensionales como son los modelos de estrella o de copo de nieve. (Kimball & Ross, 2013)

Dimensión: En data warehousing, son tablas de base de datos que contienen la información descriptiva de los entes y parámetros que intervienen en el proceso de negocio. (Kimball & Ross, 2013)

MDB: Multidimensional Database o Base de Datos Multidimensional, es un modelo que permite la identificación de los procesos de negocio y sus involucrados. Se diferencia del modelo entidad relación en la manera de organizar los datos provenientes de las fuentes. (Kimball & Ross, 2013)

OLAP: On-line Analytical Processing o Proceso Analítico en Línea, es el resultado de la organización del modelo multidimensional de una base de datos y está definido en base a las necesidades del negocio y permiten el procesamiento de grandes cantidades de datos. (Kimball & Ross, 2013)

UTS: La unidad de titulación de Sistemas de la Universidad Politécnica Salesiana. Es el ente encargado de organizar el proceso de titulación de los estudiantes que cumplan con los requisitos para obtener el título de tercer nivel. (Consejo superior de la Universidad Politécnica Salesiana, 2019)

OCR: Optical Character Recognition o Reconocimiento Óptico de Caracteres. “Es un proceso que consiste en la identificación automatizada de símbolos o caracteres pertenecientes a un determinado alfabeto, a partir de una imagen recogida mediante la lectura óptica de un texto grabado en un apoyo real” (Navarro, 2010)

REFERENCIAS BIBLIOGRÁFICAS

- Bucher, T., & Gericke, A. (2009). Process-centric business intelligence. *Business Process Management Journal*.
- Consejo superior de la Universidad Politécnica Salesiana. (2019). Normativa interna de la UPS: Unidad de Titulación Especial. En C. s. Salesiana, *Normativa interna de la UPS*. Cuenca.
- Consejo Superior UPS. (2018). *REGLAMENTO INTERNO DE RÉGIMEN ACADÉMICO DE LA UNIVERSIDAD POLITÉCNICA SALESIANA*. Cuenca.
- Han, J., Kamber, M., & Jian, P. (2012). *Data Mining Concepts and Techniques*. Waltham: Elsevier.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit*. Indianapolis: John Wiley & Sons, Inc.
- Martinez, L. (2011). Minería de Datos: cómo hallar una aguja en un pajar. *Revista Ciencia*, 18-28.
- Mera, P. J. (2016). Análisis del proceso de pruebas de calidad de software. *Ingeniería solidaria*, 163.
- Navarro, J. (2010). OCR: Sistemas de Reconocimiento Óptico de Caracteres. *Revista del Instituto Tecnológico de Informática*, 9-11.
- Pradel, J., & Raya, J. (2015). *Introducción a la ingeniería del software*. UOC.
- Shollo, A. (2013). *The role of business intelligence in organizational decision-making*. Copenague: Copenhagen Business School.

Unidad de Titulación. (2 de Diciembre de 2019). *BLOG DE LA UNIDAD DE TITULACIÓN SISTEMAS (UTES) UPS UIO*. Obtenido de titulacionsistemasups.blogspot.es: <http://titulacionsistemasups.blogspot.es>

Vicerrectorado UPS. (25 de julio de 2014). INSTRUCTIVO PARA LA ESTRUCTURA Y DESARROLLO DEL TRABAJO DE TITULACIÓN DE GRADO DE LA UNIVERSIDAD POLITÉCNICA SALESIANA. *INSTRUCTIVO PARA LA ESTRUCTURA Y DESARROLLO DEL TRABAJO DE TITULACIÓN DE GRADO DE LA UNIVERSIDAD POLITÉCNICA SALESIANA*. Cuenca, Azuay, Ecuador.

Vucevic, D., & Yaddow, W. (2012). *Testing the Data Warehouse Practicum*. Trafford.

White, S., & Miers, D. (2016). *BPMN: Guía de referencia y modelado*. Future Strategies.

Wixom, B., & Watson, H. (2010). *The Bi-Based organization*. Hershey: IGI PUBLISHING.

Zaiane, O. (1999). Principles of Knowledge Discovery in Databases. 1-15.

ANEXOS

Para revisar los anexos de este trabajo, por favor diríjase al CD.

Anexo 1: Diagramas bpmn del proceso de titulación (bpmn).

Anexo 2: Matriz de granularidad.

Anexo 3: Diccionario de datos de la base de datos de la unidad de titulación.

Anexo 4: Plan de pruebas.

Anexo 5: Manuales de usuario.

Anexo 6: Formato de inscripción de estudiantes.

Anexo 7: Formato de inscripción de estudiantes de actualización.

Anexo 8: Formato de ingreso de temas propuestos de proyectos.

Anexo 9: Formato de asignación de tutor y lector.