

UNIVERSIDAD NACIONAL DE LOJA

**ÁREA DE LA ENERGÍA, LAS INDUSTRIAS
Y LOS RECURSOS NATURALES NO
RENOVABLES**

Ingeniería en Sistemas

TÍTULO

**“COMPONENTES DE INTELIGENCIA DE NEGOCIOS EN
EL ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS
RECURSOS NATURALES NO RENOVABLES”**

**Tesis previa a la Obtención del
título de Ingeniero en Sistemas.**

Autor

Edgar Iván Tuza Cuenca

Director

**Ing. Henry Patricio Paz Arias,
Mg. Sc**

**Loja – Ecuador
2015**





CERTIFICACIÓN DEL DIRECTOR.

Ing. Henry Patricio Paz Arias, Mg. Sc.

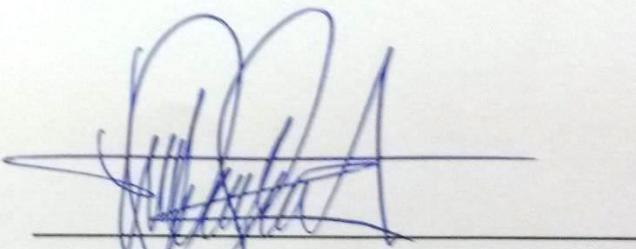
DOCENTE DE LA CARRERA DE INGENIERÍA EN SISTEMAS, DEL ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES DE LA UNIVERSIDAD NACIONAL DE LOJA.

CERTIFICA:

Que el egresado Edgar Iván Tuza Cuenca, autor del presente trabajo de titulación, el mismo que versa sobre **“COMPONENTES DE INTELIGENCIA DE NEGOCIOS EN EL ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES”** ha sido realizado bajo mi asesoramiento.

En vista de que el mismo reúne los requisitos necesarios, autorizo su presentación y defensa ante el tribunal que se designe para el efecto.

Loja, 6 de Abril del 2015.



Ing. Henry Patricio Paz Arias M.C.C.

DIRECTOR DEL PFC.



AUTORÍA.

Yo **EDGAR IVÁN TUZA CUENCA**, declaro ser autor del presente trabajo de tesis y eximo expresamente a la Universidad Nacional de Loja y a sus representantes jurídicos de posibles reclamos o acciones legales por el contenido de la misma.

Adicionalmente acepto y autorizo a la Universidad Nacional de Loja, la publicación de mi tesis en el Repositorio Institucional – Biblioteca Virtual.

Firma:

Cedula: 1104029556

Fecha: 3 de junio de 2015.



CARTA DE AUTORIZACIÓN DE TESIS POR PARTE DEL AUTOR, PARA LA CONSULTA, REPRODUCCIÓN, PARCIAL O TOTAL Y PUBLICACIÓN ELECTRÓNICA DEL TEXTO TOTAL.

Yo: **EDGAR IVÁN TUZA CUENCA**, declaro ser autor de la tesis titulada: **COMPONENTES DE INTELIGENCIA DE NEGOCIOS EN EL ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES**, como requisito para optar al grado de: **INGENIERO EN SISTEMAS**; autorizo al Sistema Bibliotecario de la Universidad Nacional de Loja para que con fines académicos, muestre al mundo la producción intelectual de la Universidad, a través de la visibilidad de su contenido de la siguiente manera en el Repositorio Digital Institucional:

Los usuarios puedan consultar el contenido de este trabajo en el RDI, en las redes de información del país y del exterior, con las cuales tenga convenio la Universidad.

La Universidad Nacional de Loja, no se responsabiliza por el plagio o copia de la tesis que realice un tercero.

Para constancia de esta autorización, en la ciudad de Loja, veinte y nueve días del mes de mayo del dos mil quince.

Firma:

Autor:

Edgar Iván Tuza Cuenca

Cedula:

1104029556

Dirección:

Loja (Ciudadela Daniel Álvarez)

Correo Electrónico:

eituzac@unl.edu.ec

Teléfono:

2569169

Celular:

0981609919

DATOS COMPLEMENTARIOS:

Director de Tesis:

Ing. Henry Patricio Paz Arias, Mg. Sc

Tribunal de grado:

Ing. Marco Augusto Ocampo Carpio, Mg. Sc

Ing. Jorge Iván Tocto, Mg. Sc

Ing. Waldemar Victorino Espinoza Tituana, Mg. Sc



AGRADECIMIENTO.

“A las estrellas no se sube por caminos llanos (José Martí)”. Con mucho esfuerzo he logrado este reto en mi vida, pero también con la ayuda de muchos familiares, de allí el compromiso de exteriorizar mi más sincero agradecimiento a mi madre **MERCY ISERAE LITA CUENCA**, una mujer extraordinaria que ha estado brindándome su apoyo incondicional, y que ha sido un pilar fundamental para obtener muchos logros en mi vida.

No puedo dejar de agradecer a Mario Alberto Lima Chuquimarca, una persona extraordinaria que también me ha brindado su constante apoyo, a mis hermanos: Danila, Omar, Pedro, Manuel, Jaqueline, Mario, que también han estado brindándome su apoyo, su constante cariño y amor, los mismos que me han servido de camino e inspiración durante mi formación como persona y como profesional.

A la Universidad Nacional de Loja, donde me forme como profesional, a todos los compañeros y amigos, profesores que impartieron sus conocimientos durante mi proceso de formación profesional, a todos ellos mi sincero agradecimiento.

El Autor.



DEDICATORIA.

Mi dedicación especial por este logro a mi querida madre, que se ha convertido en un pilar fundamental en mi vida y me ha sabido aconsejar y guiar por buenos senderos, pero sobre todo porque me ha brindado ese apoyo incondicional para cumplir mi más grande sueño de llegar a ser profesional. Por todo ese esfuerzo y sacrificio que hiciste, los mismos que me han servido como fuente de inspiración, con mucho gusto mi especial dedicatoria para ti, **MERCY ISRAELITA CUENCA**, la reina y dueña de mis ilusiones.

A mi padre **Manuel Isaías Tuza**, que aunque no esté físicamente conmigo, lo está en mi alma y en mi mente, en todos esos valores que cuando niño el me inculco y que me han servido para formarme como persona, he aquí tu hijo hecho un profesional.

Danila, Omar, Pedro, Manuel, Jakeline, Mario para ustedes mis queridos hermanos mi especial dedicación, por todo ese apoyo incondicional que me brindaron y porque siempre estuvieron pendientes para que yo cumpla con esta meta.

Edgar Iván Tuza Cuenca.



CESIÓN DE DERECHOS.

Yo **Edgar Iván Tuza Cuenca**, autor intelectual del presente Proyecto de Fin de Carrera, autorizo a la Universidad Nacional de Loja, al Área de la Energía, las Industrias y los Recursos Naturales no Renovables, y específicamente a la carrera de Ingeniería en Sistemas, el total acceso a su contenido en lo que consideren necesario.

Edgar Iván Tuza Cuenca.

1104029556



A. TÍTULO.

**“COMPONENTES DE INTELIGENCIA DE NEGOCIOS EN EL
ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS
NATURALES NO RENOVABLES”.**



B. RESUMEN.

La información se ha convertido en un elemento necesario para el surgimiento de las empresas modernas, se necesita tener conocimiento sobre sus competidores y el mercado en el que se desenvuelven, por tal motivo los datos históricos se convierten en una fuente de información valiosa, que al ser tratada adecuadamente incrementaría la competitividad de cualquier organización.

Las instituciones de educación superior, también pueden hacer uso de las grandes cantidades de información provenientes de las diferentes áreas o facultades y sus respectivas carreras, esta información puede llegar a ser de gran utilidad para descubrir posibles falencias en el desenvolvimiento académico.

El presente proyecto de investigación, describe un proceso orientado a la incorporación de componentes de inteligencia de negocios (Business Intelligence) en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables de la Universidad Nacional de Loja. Para este fin, se construyó un Data Warehouse centrado en la parte académica de la misma.

El Data Warehouse del área estará formado por dos Data Marts (estudiantes-docentes), para su construcción se realizó actividades tales como: la obtención de requerimientos del negocio (identificación de preguntas), identificación de perspectivas de análisis e indicadores de rendimiento, el análisis de la información contenida dentro del SGAWebServices, que es el principal OLTP para el desarrollo del presente proyecto, y el desarrollo de un modelo relacional que actuará como almacenamiento temporal, y en base a este último poblar los modelos dimensionales (Data Marts), basándonos en el esquema estrella y la arquitectura de Ralph Kimball.

La construcción e integración de los Data Marts, se realizó mediante procesos de Extracción, Transformación y Carga (ETL) a partir de datos contenidos en el SGAWebServices de la Universidad Nacional de Loja, los mismos que serán llevados a un almacenamiento temporal denominado STAGE_ENERGIA, para luego ser llevados a cada Data Mart y finalmente mostrar al usuario la información requerida a través de herramientas de Procesamiento Analítico en Línea (OLAP), complementándose además con herramientas de reporte, en base a los datos contenidos en cada Data Mart.



SUMMARY.

Information has become a necessary element for the emergence of modern business, you need to have knowledge about their competitors and the market in which they operate, as such historical data becomes a source of valuable information to be treated appropriately increase the competitiveness of any organization. Institutions of higher education can also make use of large amounts of information from different areas or departments and their careers, this information can be useful for potential weaknesses in academic performance.

The present investigation project describes a process aimed at incorporating components of business intelligence (BI) in the Area of Energy, Process Industries and non-renewable National University of Loja Natural Resources. So, it will build a Data Warehouse focused on the academic side of it. The Data Warehouse area will consist of two Data Marts (student-teacher), for construction activities such as performed: obtaining business requirements (identification of questions), identifying perspectives of analysis and performance indicators, analysis of the information contained within SGAWebServices which is the main OLTP development of this project and the development of a relational model that will act as temporary storage, and based on the latter populate dimensional models (Data Marts), based on the star schema and following Ralph Kimball's architecture. The implementation and integration of Data Marts was performed by processes Extraction, Transformation and Loading (ETL) from data contained in the SGAWebServices of the National University of Loja, the same that will be taken to a temporary storage called STAGE_ENERGIA, then be brought to each Data Mart, and finally show the user the information required by tools Online Analytical Processing (OLAP), and complemented with additional reporting tools, based on the data contents in each Data Mart.



ÍNDICE DE CONTENIDOS.

CERTIFICACIÓN DEL DIRECTOR.....	I
AUTORÍA.....	II
CARTA DE AUTORIZACIÓN.....	III
AGRADECIMIENTO.....	IV
DEDICATORIA.....	V
CESIÓN DE DERECHOS.....	VI
A. TÍTULO.....	VII
B. RESUMEN.....	VIII
SUMMARY.....	IX
ÍNDICE DE CONTENIDOS.....	X
ÍNDICE DE FIGURAS.....	XVI
ÍNDICE DE TABLAS.....	XIX
C. INTRODUCCIÓN.....	1
D. REVISIÓN DE LITERATURA.....	3
CAPÍTULO I: INTELIGENCIA DE NEGOCIOS.....	3
1.1. Introducción a la Inteligencia de Negocios.....	3
1.2. ¿Qué es Inteligencia de Negocios?.....	3
1.3. Arquitecturas de Inteligencia de Negocios.....	4
1.3.1. Corporate Information Factory (CIF).....	4
1.3.2. Arquitectura Multidimensional.....	5
1.3.3. Arquitectura CIF Vs Arquitectura MD.....	6
CAPÍTULO II: DATA WAREHOUSE.....	7
2.1. ¿Qué es Data Warehouse?.....	7
2.2. Arquitectura de un Data Warehouse.....	7
2.2.1. Arquitectura de flujo de datos.....	7



2.2.1.1.	Arquitectura de flujo de fatos general.....	7
2.2.1.2.	Arquitectura de flujo de datos simple con un almacén de datos dimensional (Simple DDS).....	8
2.2.2.	Arquitectura del sistema.	9
CAPITULO III DATA WAREHOUSING.....		10
3.1.	Data Warehousing.	10
3.2.	Arquitectura Data Warehousing.	10
3.2.1.	OLTP.....	10
3.2.2.	Load Manager.	11
3.2.3.	Data Warehouse Manager.....	11
3.2.4.	Query Manager.....	11
3.2.5.	Herramientas de consulta y análisis.	11
3.2.6.	Usuarios.	11
CAPITULO IV: BASE DE DATOS MULTIDIMENSIONALES.....		12
4.1.	Base de datos Multidimensional.....	12
4.1.1.	Tablas de dimensiones.....	13
4.1.1.1.	Tabla de dimensión tiempo.	14
4.1.1.2.	Dimensiones lentamente cambiantes con el tiempo.....	14
4.1.2.	Tablas de hechos.	15
4.1.2.1.	Indicadores.	16
4.1.2.2.	Atributos.....	16
4.1.2.3.	Jerarquías.....	16
CAPITULO V: HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS.....		17
5.1.	Herramientas de inteligencia de negocios en el mercado.....	17
5.2.	Plataformas Open Source enfocadas a Inteligencia de Negocios.	18
CAPITULO XI: OLAP (On Line Analytical Processing).		20
6.1.	Sistemas OLAP (On Line Analytical Processing).....	20
6.2.	MOLAP (OLAP Multidimensional).	20



E.	MATERIALES Y MÉTODOS.....	22
F.	RESULTADOS.....	24
1.	Recopilación de Información.....	24
1.1.	Analizar y comprender la Naturaleza del Negocio.....	24
1.2.	Obtener los Requerimientos e Indicadores claves de rendimiento.....	24
1.3.	Analizar las distintas fuentes de datos del negocio.....	26
2.	Selección de la Metodología.....	26
3.	Desarrollo del Data Warehouse aplicando la metodología HEFESTO.....	28
3.1.	Modelo lógico y físico del Data Warehouse.....	28
3.1.1.	Etapa 1: Análisis de Requerimientos.....	31
3.1.1.1.	Identificar indicadores y perspectivas de análisis.....	31
3.1.1.2.	Modelo conceptual.....	32
3.1.2.	Etapa 2: Análisis de los OLTP.....	33
3.1.2.1.	Construcción de los indicadores.....	35
3.1.2.2.	Establecer Correspondencias.....	36
3.1.2.3.	Nivel de granularidad.....	39
3.1.2.4.	Modelo conceptual ampliado.....	42
3.1.3.	Etapa 3: Modelo lógico de los Data Marts.....	43
3.1.3.1.	Tablas de hechos y tablas de dimensiones.....	45
3.1.4.	Etapa 4: Integración de datos (Procesos ETL).....	45
3.1.4.1.	Arquitectura de flujo de datos.....	45
3.1.4.2.	Implementación de flujos de carga.....	49
3.1.4.2.1.	Flujo de trabajo STAGE_ENERGIA.....	50
3.1.4.2.2.	Caracterización de datos.....	51
3.1.4.2.3.	Flujo de trabajo UNIR_TRANSF_ESTUDIANTE.....	53
3.1.4.2.4.	Flujo de trabajo UNIR_TRANSF_DOCENTE.....	53
3.1.4.2.5.	Flujo de trabajo UNIR_DM_STAGE.....	54
3.1.4.2.6.	Actualización del Data Warehouse.....	54



3.2.	Aplicación del DW para el análisis de la información académica del AEIRNNR (Extracción de la información).....	55
3.2.1.	Creación de cubos multidimensionales (Reportes Dinámicos).....	55
3.2.1.1.	Cubo Docentes.	56
3.2.1.2.	Cubo Estudiantes.....	57
3.2.2.	Generación de reportes personalizados.	58
3.2.3.	Presentación de información al usuario final.....	58
3.2.3.1.	Tasa de Aprobados y Reprobados.....	59
3.2.3.2.	Tasa de profesores por Categoría.....	60
3.2.3.3.	Formación Posgrado Docentes.....	61
3.2.3.4.	Total de docentes por Oferta Académica y Carrera.....	62
3.2.3.5.	Promedio de notas por Oferta Académica y Carrera.....	64
3.2.3.6.	Tasa de estudiantes por Oferta Académica – Distribución estudiantes por Carreras.	65
G.	DISCUSIÓN.....	66
1.	DESARROLLO DE LA PROPUESTA ALTERNATIVA.....	66
2.	VALORACIÓN TÉCNICA ECONÓMICA AMBIENTAL.	68
H.	CONCLUSIONES.	72
I.	RECOMENDACIONES.....	73
J.	BIBLIOGRAFÍA.	74
K.	ANEXOS.....	78
	Anexo 1: Anteproyecto del Trabajo de Titulación.	78
A.	Tema.....	82
B.	Problemática.	83
1.	Situación Problemática.....	83
2.	Problema de Investigación.	84
C.	Justificación.	85
D.	Objetivos.....	87



1. Objetivo General.....	87
2. Objetivos Específicos.....	87
E. Alcance.....	88
F. Metodología.....	89
G. Cronograma.....	90
H. Presupuesto.....	91
I. Bibliografía.....	93
J. Anexos.....	94
Anexo 2: Modelo de Entrevista.....	95
Anexo 3: Diccionario de datos.....	97
1. Dimensión DIM_CARRERA.....	97
2. Dimensión DIM_TITULO.....	97
3. Dimensión DIM_CATEGORIA.....	98
4. Dimensión DIM_GENERO.....	99
5. Dimensión DIM_FECHA.....	99
6. Dimensión DIM_ESTUDIANTE.....	100
7. Dimensión DIM_DOCENTE.....	100
8. Tabla de hechos FACT_ANALISIS_RENDIMIENTO.....	101
9. Tabla de hechos FACT_ANALISIS_DOCENTE.....	102
Anexo 4: Conexión base de datos con Pentaho Data Integration (PDI).....	103
Anexo 5: ETL CARGA_OFERTAS_STAGE.....	104
Anexo 6: Configuración de transformaciones del trabajo UNIR_TRANSF_ESTUDIANTE.....	109
1. ETL CARGA_DIM_CARRERA_ESTUDIANTE.....	109
2. ETL DIM_FECHA_ESTUDIANTE.....	111
3. ETL HECHOS_ESTUDIANTE.....	111
Anexo 7: Reportes con Pentaho Report Designer.....	116
Anexo 8: Creación de cubos OLAP con Pentaho Schema Workbench.....	122



Anexo 9: Estructura Académica del AEIRNNR.....	126
Anexo 10: Pentaho BIServer.....	128
Anexo 11: Certificación Summary.	131
Anexo 12: Certificación Revisión Literaria.	132
Anexo 13: Artículo Científico.	133



ÍNDICE DE FIGURAS.

Figura 1 Arquitectura CIF.....	4
Figura 2 Arquitectura Multidimensional (MD).	5
Figura 3 Arquitectura de Flujo de Datos General.	7
Figura 4 Arquitectura simple DDS.....	8
Figura 5 Ejemplo de Arquitectura de un Sistema DW.	9
Figura 6 Arquitectura Data Warehousing.	10
Figura 7 Esquema en Estrella.....	12
Figura 8 Esquema Copo de Nieve.	13
Figura 9 Esquema Constelación o Copo de Nieve.....	13
Figura 10 Tabla de hechos.	15
Figura 11 Jerarquía dimensión tiempo.....	16
Figura 12 Cuadrante Mágico de Gartner de Plataformas Business Intelligence.	17
Figura 13 Funcionamiento de los modelos MOLAP.	20
Figura 14 Modelo de Componentes de Inteligencia de Negocios para el AEIRNNR. ...	30
Figura 15 Modelo conceptual DM_ESTUDIANTES_AEIRNNR.....	32
Figura 16 Modelo Conceptual DM_DOCENTES_AEIRNNR.	32
Figura 17 Correspondencias entre métodos del SGASWebServices y el modelo conceptual DM_ESTUDIANTE_AEIRNNR.....	37
Figura 18 Correspondencias entre métodos del SGASWebServices y el modelo conceptual DM_DOCENTES_AEIRNNR.	38
Figura 19 Modelo conceptual ampliado DM_ESTUDIANTES_AEIRNNR.	42
Figura 20 Modelo conceptual ampliado DM_DOCENTES_AEIRNNR.....	43
Figura 21 Modelo de datos dimensional DM_ESTUDIANTES_AEIRNNR.	44
Figura 22 Modelo de datos dimensional DM_DOCENTES_AEIRNNR.....	44
Figura 23 Arquitectura de flujo de datos con PDI.	45
Figura 24 Salida de información SGASWebServices.	46
Figura 25 Modelo Relacional STAGE_ENERGIA.....	47
Figura 26 Ejemplo de Transformación en PDI.	48
Figura 27 Ejemplo de trabajo (Job) con PDI.	48
Figura 28 Job STAGE_ENERGIA.....	50
Figura 29 Error detectado mediante Caracterización de datos.....	52
Figura 30 Job UNIR_TRANSF_ESTUDIANTE.....	53



Figura 31 Job UNIR_TRANSF_DOCENTE.....	54
Figura 32 Job UNIR_DM_STAGE.....	54
Figura 33 Representación CUBO_PROFESORES.....	56
Figura 34 Representación CUBO_RENDIMIENTO.....	57
Figura 35 Estudiantes Aprobados y Reprobados por carrera.....	59
Figura 36 Tasa de profesores por Categoría Laboral.....	60
Figura 37 Formación posgrado docente.	61
Figura 38 Docentes por Oferta Académica.	62
Figura 39 Distribución docente por Oferta Académica y Carrera.	63
Figura 40 Promedio de notas por Oferta Académica y Carrera.....	64
Figura 41 Tasa de estudiantes por Oferta Académica – Distribución estudiantes por Carrera.	65
Figura 42 Cronograma del Proyecto de Titulación.	90
Figura 43 Conexión a BDD con PDI.....	103
Figura 44 Transformación CARGA_OFERTAS_STAGE.....	104
Figura 45 Configuración del paso PERIODOS_STAGE.....	104
Figura 46 Configuración del paso WS_OFERTAS.....	105
Figura 47 Salida del paso WS_OFERTAS.....	106
Figura 48 Configuración del paso JSON1.....	106
Figura 49 Salida del paso JSON1.....	107
Figura 50 Configuración del paso JSON2.....	107
Figura 51 Salida del paso JSON2.....	108
Figura 52 Configuración del paso INSERTAR/ ACTUALIZAR (Ofertas).....	108
Figura 53 Transformación CARGA_DIM_CARRERA_ESTUDIANTE.	109
Figura 54 Configuración del paso DATOS_CARRERA.....	109
Figura 55 Configuración del paso DIM_CARRERA.....	110
Figura 56 Transformación DIM_FECHA_ESTUDIANTE.	111
Figura 57 Configuración del paso OFERTAS_PERIODOS.....	111
Figura 58 Transformación HECHOS_ESTUDIANTE.	112
Figura 59 Configuración del paso UNION_STAGE.....	112
Figura 60 Código JavaScript empleado en el salto CONVERSION.....	112
Figura 61 Configuración del paso MAPEO_APREP.....	113
Figura 62 Configuración del paso DIM_CARRERA.....	114
Figura 63 Configuración del paso INSERTAR/ACTUALIZAR_HECHOS.....	115
Figura 64 Generar reporte con Report Wizard.....	116



Figura 65 Plantillas Report Wizard.....	117
Figura 66 Conexión a base de datos con Report Wizard.	117
Figura 67 Configuración de acceso a base de datos con Pentaho Report Designer.	118
Figura 68 Consulta SQL ofertas.....	118
Figura 69 Consulta SQL formación_posgrado.	119
Figura 70 Consulta SQL carreras.	119
Figura 71 Creación de parámetros en Report Designer.	119
Figura 72 Configuración de parámetros con Report Designer.....	120
Figura 73 Publicar reporte con Report Designer.	120
Figura 74 Agregar ruta de ubicación del archivo con PRD.....	121
Figura 75 Confirmación de publicación de reporte con PRD.....	121
Figura 76 Acceso a base de datos con PSW.....	122
Figura 77 Creación de un esquema con PSW.	122
Figura 78 Creación de un cubo con PSW.	123
Figura 79 Agregar tabla de hechos a un cubo con PSW.....	123
Figura 80 Crear dimensión en un cubo con PSW.	123
Figura 81 Agregar dimensión a un cubo con PSW.....	124
Figura 82 Agregar campos a mostrar en una jerarquía con PSW.	124
Figura 83 Cálculo de indicadores con PSW.....	124
Figura 84 Publicar cubo con Pentaho Schema Workbench.	125
Figura 85 Estructura académica del AEIRNNR.....	126
Figura 86 Consola de login de Pentaho BIServer.	128
Figura 87 Consola de Administración de Pentaho BIServer.....	129
Figura 88 Conexión a BDD con Pentaho BIServer.....	129
Figura 89 Configuración de BDD con Pentaho BIServer.....	130



ÍNDICE DE TABLAS.

TABLA I COMPARATIVA ARQUITECTURA CIF Y MD.	6
TABLA II PLATAFORMAS OPEN SOURCE ENFOCADAS A INTELIGENCIA DE NEGOCIOS.	19
TABLA III VENTAJAS Y DESVENTAJAS DE MOLAP.	21
TABLA IV IDENTIFICACIÓN DE REQUERIMIENTOS.	25
TABLA V METODOLOGÍAS PARA INTELIGENCIA DE NEGOCIOS.	27
TABLA VI INDICADORES Y PERSPECTIVAS DE ANÁLISIS.	31
TABLA VII MÉTODOS SGAWEBSERVICES INFORMACIÓN INSTITUCIONAL.....	33
TABLA VIII MÉTODOS SGAWEBSERVICES INFORMACIÓN PERSONAL.	33
TABLA IX MÉTODOS SGAWEBSERVICES INFORMACIÓN ACADÉMICA.....	34
TABLA X SALIDA MÉTODO SGAWS_DATOS_ESTUDIANTE.	39
TABLA XI SALIDA MÉTODO SGAWS_CARRERAS_AREA.....	39
TABLA XII SALIDA MÉTODO SGAWS_DATOS_DOCENTE.	40
TABLA XIII SALIDA MÉTODO SGAWS_PERIODOS_LLECTIVOS.....	41
TABLA XIV SALIDA MÉTODO SGAWS_OFERTAS_ACADEMICAS.	41
TABLA XV RECURSOS HUMANOS.	68
TABLA XVI SERVICIOS BÁSICOS.....	68
TABLA XVII RECURSOS MATERIALES.	69
TABLA XVIII RECURSOS HARDWARE.....	69
TABLA XIX RECURSOS SOFTWARE.	70
TABLA XX RECURSOS TÉCNICOS Y TECNOLÓGICOS.	70
TABLA XXI COSTO DEL PROYECTO (APROXIMACIÓN).	70
TABLA XXII TALENTO HUMANO PARA EL DESARROLLO DEL PROYECTO.	91
TABLA XXIII RECURSOS TÉCNICOS PARA EL DESARROLLO DE PROYECTO....	91
TABLA XXIV SERVICIOS BÁSICOS PARA EL DESARROLLO DE PROYECTO.....	91
TABLA XXV RECURSOS FÍSICOS PARA EL DESARROLLO DE PROYECTO.	92
TABLA XXVI PRESUPUESTO GENERAL PARA EL DESARROLLO DEL PROYECTO.	92
TABLA XXVII DIMENSIÓN DIM_CARRERA.	97
TABLA XXVIII DIMENSIÓN DIM_TITULO.	98
TABLA XXIX DIMENSIÓN DIM_CATEGORIA.....	98
TABLA XXX DIMENSIÓN DIM_GENERO.	99
TABLA XXXI DIMENSIÓN DIM_FECHA.....	99



TABLA XXXII DIMENSIÓN DIM_ESTUDIANTE.	100
TABLA XXXIII DIMENSIÓN DIM_DOCENTE.	101
TABLA XXXIV TABLA DE HECHOS FACT_ANALISIS_RENDIMIENTO.	101
TABLA XXXV TABLA DE HECHOS FACT_ANALISIS_DOCENTE.	102



C. INTRODUCCIÓN.

En la actualidad todas las organizaciones están comprendiendo la importancia de la gestión de la información y las ventajas que acarrea consigo su adecuada utilización, este proceso de gestión consiste en dar un buen uso a los distintos tipos de datos de la empresa, por tal motivo se busca incorporar aplicaciones que abarquen todas las actividades enfocadas al análisis del negocio.

El Área de la Energía, las Industrias y los Recursos Naturales no Renovables (AEIRNNR) de la Universidad Nacional de Loja, es un claro ejemplo de esta necesidad, su alto grado de información en cuanto a la parte académica se refiere, ha generado nuevas necesidades a satisfacer desde la perspectiva de la gestión, en donde el ámbito de interés lo constituyen la parte académica y la docencia impartida dentro del área.

En el presente proyecto de titulación denominado “**Componentes de Inteligencia de Negocios en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables**”, se ha tomado como caso de estudio el área antes mencionada de la Universidad Nacional de Loja, construyéndose dos Data Marts (docentes, estudiantes) para el análisis de la información académica de la misma. Cada uno de los Data Marts desarrollados serán poblados a partir de datos académicos contenidos en el SGASWebServices de la Universidad Nacional de Loja.

La arquitectura de inteligencia de negocios en la que se ha puesto énfasis para el desarrollo del presente proyecto es la propuesta por Ralph Kimball, considerado como uno de los padres del Data Warehousing, además se ha enfocado el proyecto a la utilización de la metodología HEFESTO y la plataforma Open Source para Inteligencia de Negocios (Pentaho Business Intelligence).

La plataforma Pentaho Business Intelligence se adapta a la arquitectura antes mencionada, y sirve de apoyo en todos y cada uno de los procesos de un sistema Data Warehousing, que van desde la obtención de datos de entrada (Pentaho Data Integration), hasta la entrega de información de salida (Report Designer, Schema Worbench, etc.).

El proyecto de titulación se lo ha estructurado en dos partes. En primera instancia se abarca conceptos relacionados con inteligencia de negocios, además se explica y describe la arquitectura propuesta por Ralph Kimball y Bill Inmon, realizando una



comparativa entre ambas arquitecturas. Otro de los aspectos importantes que abarca este capítulo es la diferencia que existe entre Data Warehouse y Data Warehousing, las arquitecturas de flujos de datos existentes para los sistemas Data Warehouse, definiendo además cada uno de los componentes que intervienen durante un proceso Data Warehousing.

Luego de revisar conceptos relacionados con inteligencia de negocios, se abarca al que constituye el capítulo más importante, ya que se centra en la construcción de cada Data Mart, se define la arquitectura de flujo de datos adoptada y como la herramienta Open Source Pentaho Data Integration, apoyan en cada etapa del proceso de construcción de los Data Marts. En este capítulo se definen las necesidades de información que existen dentro del AEIRNNR (requerimientos), se explica el diseño de la base de datos relacional que actuará como almacenamiento temporal y los modelos dimensionales o Data Marts que conforman el Data Warehouse. Además, se explica el procedimiento para identificar perspectivas de análisis e indicadores de rendimiento, y los procesos ETL (Extracción, Transformación y Carga) que se utilizaron para poblar el almacén de datos temporal y los modelos dimensionales.

Finalmente, se realizará el modelamiento de cubos OLAP (Pentaho Schema Workbench), reportes mediante el módulo Pentaho Report Designer, para mostrarlos al usuario final a través del módulo Pentaho BI Server.



D. REVISIÓN DE LITERATURA.

CAPÍTULO I: INTELIGENCIA DE NEGOCIOS.

1.1. Introducción a la Inteligencia de Negocios.

La aparición de los ordenadores en la década de 1970 popularizó la utilización de las computadoras personales y sistemas transaccionales en los procesos administrativos de las organizaciones, llámense estas comerciales, educativas, industriales, etc. Pero dichos sistemas en su mayoría no son construidos con el fin de brindar síntesis, análisis, consolidación, búsquedas, y proyecciones en base a la información histórica que generan. Sin embargo, una de las formas de sacar el mayor provecho a este tipo de información es utilizar un conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento, mediante el análisis de datos llamada "Inteligencia de Negocios", también conocida como inteligencia empresarial o BI (del inglés Business Intelligence). El concepto de inteligencia de negocios se analiza desde tres perspectivas diferentes: Tomar decisiones rápidamente, convertir datos en información y utilizar un método razonable para la gestión empresarial [7].

1.2. ¿Qué es Inteligencia de Negocios?

La inteligencia de negocios, se define como la habilidad corporativa para tomar decisiones. Esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos, y aplicar en ellos técnicas analíticas de extracción de conocimiento, los datos pueden ser estructurados para que indiquen las características de un área de interés, generando el conocimiento sobre los problemas y oportunidades del negocio, para que puedan ser corregidos y aprovechados respectivamente [2]. Al contar con la información exacta y en tiempo real, es posible aparte de lo ya mencionado, identificar y corregir situaciones antes de que se conviertan en problemas y en potenciales pérdidas de control para la empresa, pudiendo conseguir nuevas oportunidades o readaptarse frente a la ocurrencia de sucesos inesperados [3].

Las definiciones antes expuestas pueden representarse a través de la siguiente fórmula:

Datos + Análisis = Conocimiento

Por otra parte, la inteligencia de negocios también es considerada como concepto que integra el almacenamiento y el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración [2]. Dentro de los principales tipos de análisis que nos proporciona la inteligencia de negocios tenemos:

- Análisis descriptivo (Reporteo, Procesamiento analítico en línea Cubos OLAP, etc.).
- Análisis predictivo (Minería de datos).

1.3. Arquitecturas de Inteligencia de Negocios.

Bill Inmon y Ralph Kimball, conocidos como los padres del Data Warehousing proponen su propia arquitectura, similar pero con diferentes variantes para la implantación de proyectos relacionados con inteligencia de negocios. Las dos propuestas de arquitectura, tienen el mismo propósito pero su principal diferencia radica en la manera de modelar los datos [8], [9]. A continuación se detallan cada una de las arquitecturas.

1.3.1. Corporate Information Factory (CIF).

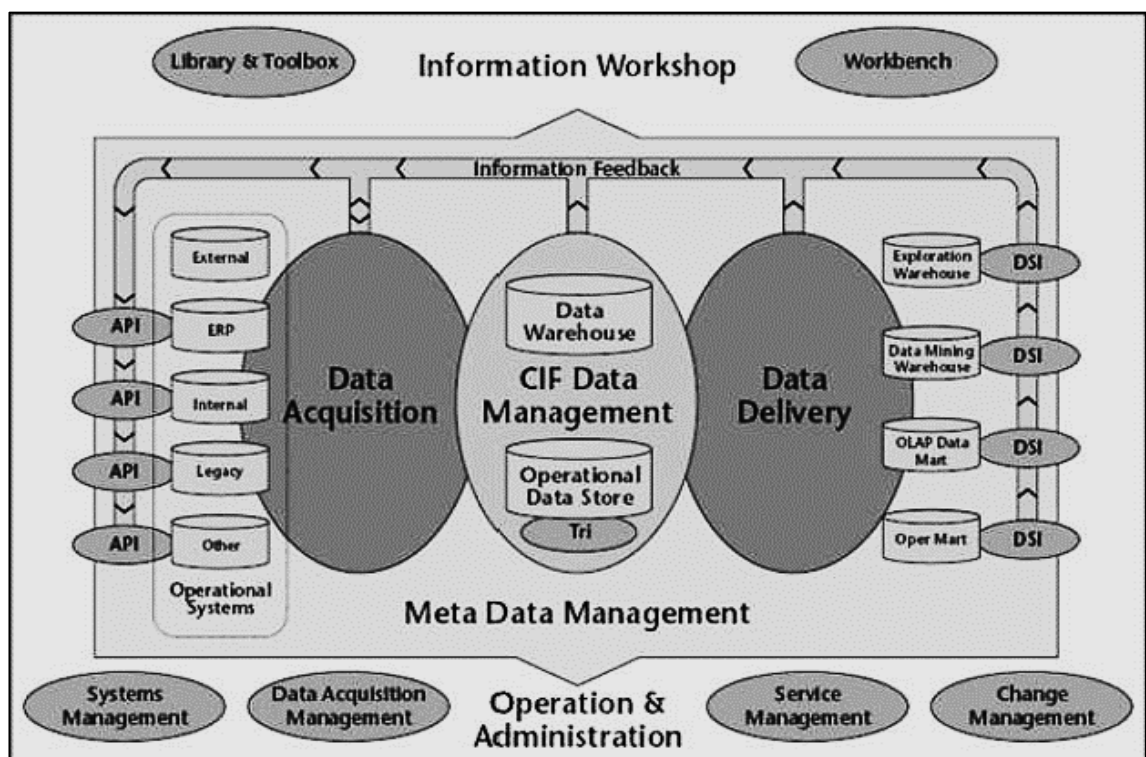


Figura 1 Arquitectura CIF.

La fábrica de información corporativa (CIF) es una arquitectura conceptual que describe y categoriza los almacenes de información usados para operar y gestionar con éxito una infraestructura de inteligencia de negocios robusta. El uso de esta arquitectura o de otras más sencillas va a depender del tipo de compañía, los requerimientos de análisis y hasta donde se quiera llegar en el uso del BI [9]. Los elementos que forman el CIF, de forma resumida, son los siguientes:

- a) **Data Warehouse:** es el punto central de la integración de datos. Centraliza toda la información, nos da una visión en común de la información de toda la organización y proporciona los datos para llenar de contenido el resto de almacenes de datos específicos.
- b) **Operational Data Store:** es un almacén de datos, como el DW, pero orientado a la toma de decisiones tácticas. Se alimenta de datos actuales de los sistemas operacionales, no es un sistema histórico y suelen ser mucho más rápidos para permitir la toma de decisiones sobre los datos de operación del negocio.
- c) **Data Acquisition:** son todas las herramientas y sistemas de gestión que nos permite la extracción, transformación y carga de los datos provenientes de los diferentes sistemas origen a nuestro Data Warehouse.
- d) **Data Delivery:** son las operaciones de agregación de la información, filtrado por dimensiones específicas o requerimientos de negocio, reformato o procesamiento de la información para soportar el uso de herramientas de inteligencia de negocios específicas, y finalmente, la transmisión de la información a través de la organización [9].

1.3.2. Arquitectura Multidimensional.

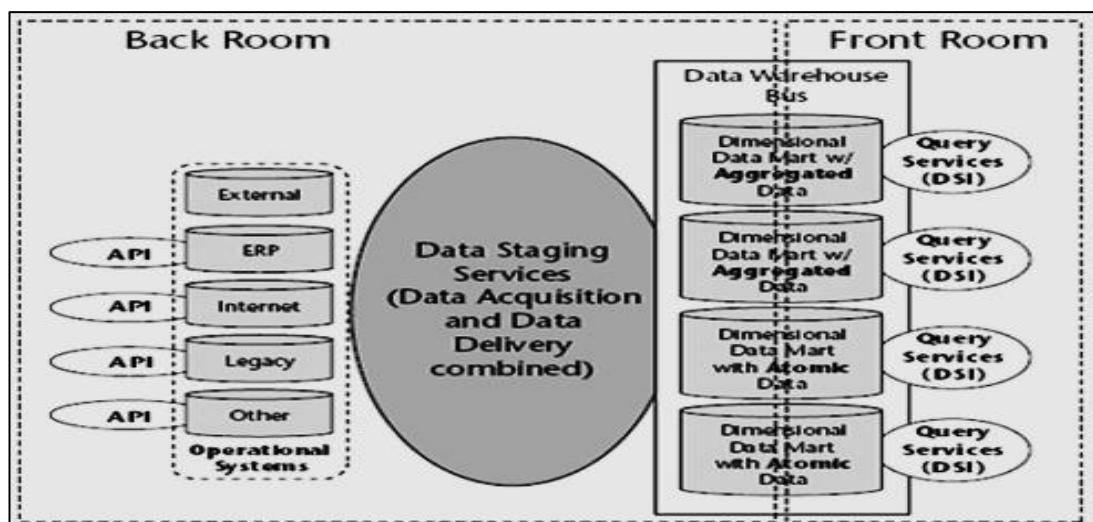


Figura 2 Arquitectura Multidimensional (MD).



La arquitectura multidimensional se basa en la premisa de que todos los tipos de análisis de BI se fundamentan en un diseño de datos multidimensional. Dentro de esta arquitectura el esquema en estrella constituye el más útil para la toma de decisiones, de allí que comienza y termina enfocada en unidades de negocio individuales con un requerimiento específico de inteligencia de negocios, es decir no se necesita crear un Data Warehouse, sino que este se compone de la integración de los Data Marts atómicos o detallados y Data Marts agregados [10].

1.3.3. Arquitectura CIF Vs Arquitectura MD.

La principal diferencia que existe entre ambas arquitecturas radica en el modelamiento de datos, mientras que la arquitectura CIF propone modelar el Data Warehouse usando un modelo relacional, la arquitectura MD propone utilizar un modelamiento de datos dimensional con el fin de construir los diferentes Data Marts. En la Tabla I se muestra de forma resumida una breve comparativa entre ambas arquitecturas (CIF - MD).

TABLA I COMPARATIVA ARQUITECTURA CIF Y MD.

Características	Arquitectura CIF	Arquitectura MD
Técnica de modelamiento de datos	Relacional Dimensional	Dimensional
Enfoque	TOP - DOWN	BOTTOM - UP
Necesita construir un DW	SI	NO
Datamarts derivados del DW	SI	NO
Vista empresarial de los datos	Si, a través de un Data Warehouse	Parcialmente a través de un Data Warehouse en Bus
Datos detallados	SI	SI
Datos sumariados	SI	SI
Libre de procesos (process - free)	SI	NO
Tiempo de implantación.	Alto	Bajo
Costo de implantación.	Alto	Bajo
Soporta almacenes de datos operacionales (ODS) y Marts Operacionales.	SI	NO

CAPÍTULO II: DATA WAREHOUSE.

2.1. ¿Qué es Data Warehouse?

Un Data Warehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence. La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales, etc.). Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales) [6], [13], [14].

2.2. Arquitectura de un Data Warehouse.

2.2.1. Arquitectura de flujo de datos.

La arquitectura de flujo de datos es una configuración de almacenes de datos dentro de un sistema de Data Warehouse, que indica como los datos fluyen desde los sistemas fuente a través de los almacenes de datos, hasta llegar a las aplicaciones usadas por el usuario final [14].

2.2.1.1. Arquitectura de flujo de datos general.

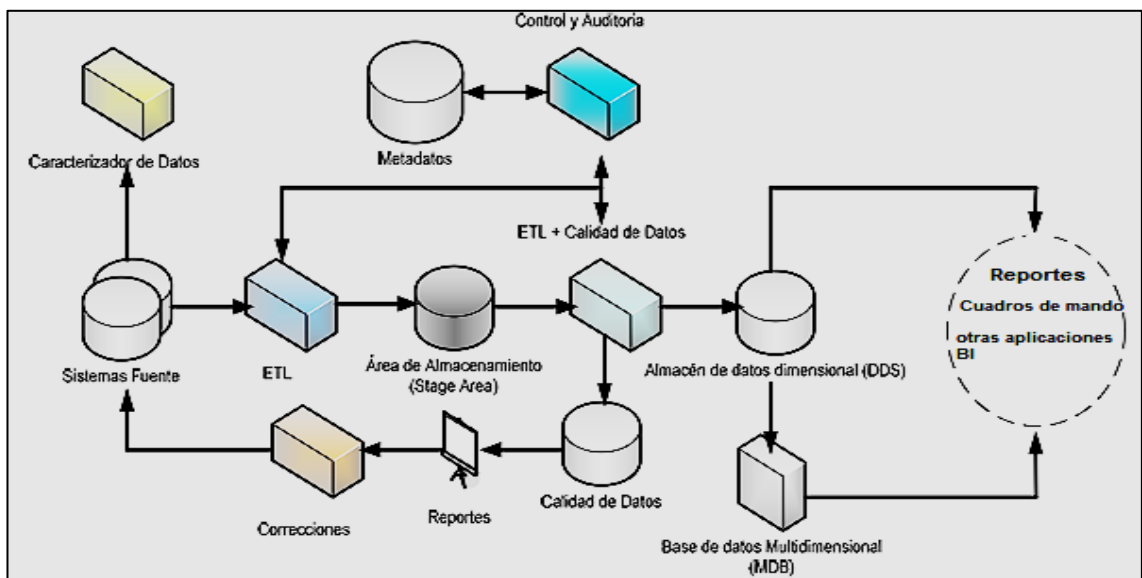


Figura 3 Arquitectura de Flujo de Datos General.

Para el desarrollo de cualquier sistema de Data Warehouse, una arquitectura que se relaciona con las propuestas de arquitecturas de inteligencia de negocios de Bill Inmon y Ralph Kimball para un sistema de Data Warehouse, es la que se muestra en la Fig. 3.

2.2.1.2. Arquitectura de flujo de datos simple con un almacén de datos dimensional (Simple DDS).

Este tipo de arquitectura es simple, es decir los datos no tienen que ser cargados en ningún tipo de almacén de datos normalizado, sino que son extraídos a un área de almacenamiento temporal denominado Stage Área, y de aquí mediante procesos ETL son llevados al almacén de datos dimensional (DDS), aplicando además reglas de calidad de datos. Dentro de esta arquitectura se puede crear más almacenes de datos dimensionales (DDS), y es recomendable utilizarla cuando se tiene una sola fuente de datos y no se necesita integrar datos provenientes de otras fuentes. Otra de las características a considerar dentro de este tipo de arquitectura, es su relación con la propuesta de Ralph Kimball, para lograr una arquitectura de inteligencia de negocios multidimensional (MD). Se debe recalcar que esta arquitectura almacena los metadatos de aplicaciones de inteligencia de negocios, así como los procesos ETL mediante componentes de control y auditoría. En la Fig. 4 se muestra la arquitectura simple con un almacén de datos dimensional [14].

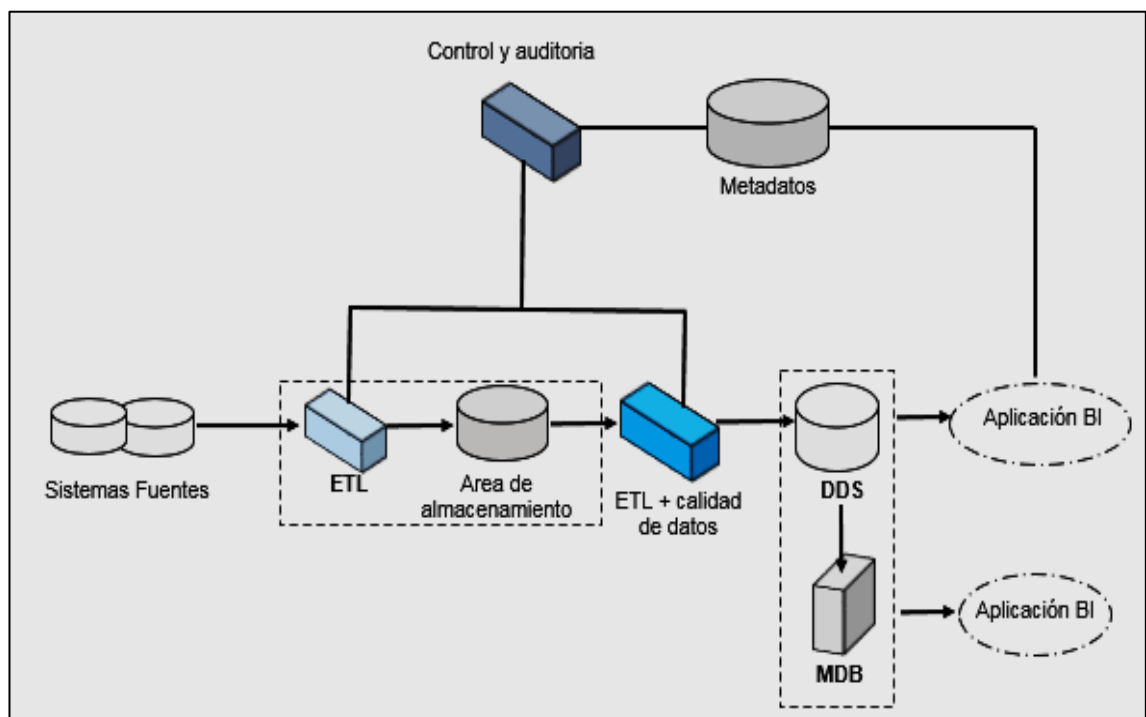


Figura 4 Arquitectura simple DDS.



2.2.2. Arquitectura del sistema.

El diseño de la arquitectura de un sistema Data Warehouse se refiere a la infraestructura física en donde el sistema se ejecutará. La arquitectura física contiene tres elementos principales que son el hardware, las redes y el almacenamiento. El hardware se refiere básicamente al conjunto de servidores que utilizará el sistema de Data Warehouse, estos pueden ser: servidores de bases de datos, servidores de aplicaciones, servidores ETL, etc.

Las redes constituyen el medio de comunicación, entre los servidores y es importante que su rendimiento y disponibilidad sean altos. La capacidad de almacenamiento es otro elemento importante en la arquitectura de un Data Warehouse [14]. Es de importancia recalcar que un arquitecto de Data Warehouse no diseña la arquitectura física del sistema sino la arquitectura de flujo de datos, pero podría ser útil si conoce sobre los temas necesarios para su diseño [14].

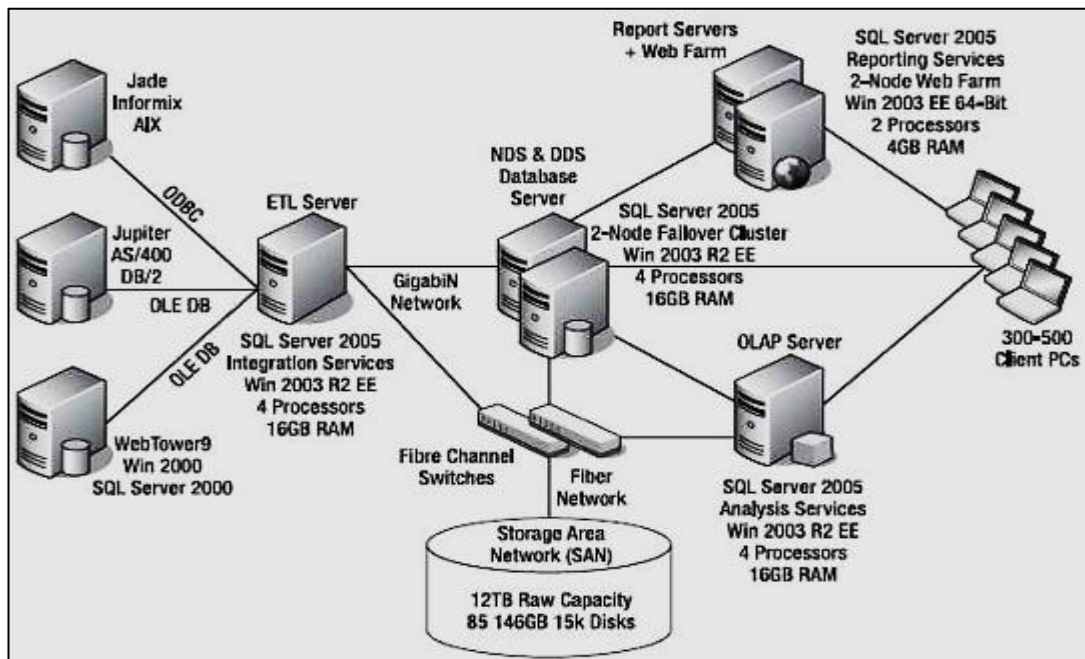


Figura 5 Ejemplo de Arquitectura de un Sistema DW.

CAPITULO III DATA WAREHOUSING.

3.1. Data Warehousing.

El Data Warehousing es el proceso que permite gestionar datos guardados en diversos formatos, fuentes y tipos, para luego depurarlos e integrarlos, además de almacenarlos en un solo destino o base de datos que permita su posterior análisis y exploración [3].

3.2. Arquitectura Data Warehousing.

Como se mencionó anteriormente Data Warehouse y Data Warehousing son dos términos diferentes, teniendo en cuenta que ya se han detallado claramente las características generales de ambos términos, se definirán y describirán todos los componentes que intervienen en un proceso Data Warehousing.

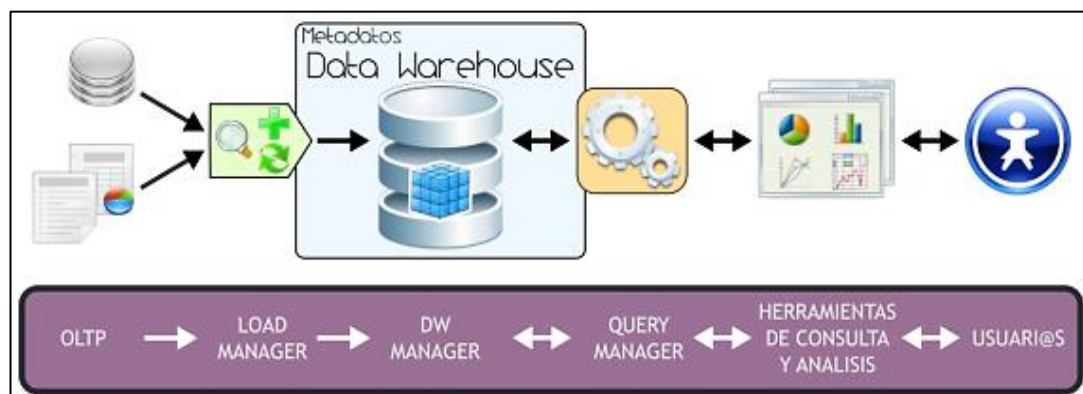


Figura 6 Arquitectura Data Warehousing.

3.2.1. OLTP.

OLTP (On Line Transaction Processing) representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer. Los paquetes de software para OLTP se basan en la arquitectura cliente servidor, ya que suelen ser utilizados por empresas mediante una red informática distribuida [3]. Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran:

- Archivos de textos.
- Web Services (OLTP que se utilizará para la realización del presente proyecto, SGA Web Services de la Universidad Nacional de Loja).
- Hipertextos, hojas de cálculo
- Informes semanales, mensuales, anuales, etc.
- Bases de datos transaccionales [1].



3.2.2. Load Manager.

Permite extraer los datos desde los OLTP, para luego manipularlos, integrarlos y transformarlos, para posteriormente cargar los resultados obtenidos al DW [3]. Los ETL (Extracción, Transformación y Carga) son los que cumplirán con tal fin, para ello es necesario contar con algún sistema que se encargue de esto (Pentaho Data Integration).

3.2.3. Data Warehouse Manager.

Permite transformar e integrar los datos fuentes y del almacenamiento intermedio en un modelo adecuado para la toma de decisiones, es decir gestiona el depósito de datos a través de tablas de hechos y tablas de dimensiones, y lo organiza en torno a una base de datos multidimensional. Esto permite que se puedan crear cubos multidimensionales u otras estructuras de datos.

3.2.4. Query Manager.

Este componente realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, tales como Join y agregaciones, y de consultas propias del análisis de datos. Query Manager recibe las consultas del usuario, las aplica a la estructura de datos correspondiente (cubo multidimensional, Business Models, etc.) y devuelve los resultados obtenidos.

3.2.5. Herramientas de consulta y análisis.

Las herramientas de consulta y análisis son sistemas que permiten al usuario realizar la exploración de datos del Data Warehouse. Básicamente constituyen el nexo entre el depósito de datos y los usuarios.

3.2.6. Usuarios.

Los usuarios que posee el DW, son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio, es por ello que se hace tanto énfasis en la integración, limpieza de datos, etc. para poder conseguir que la información posea toda la calidad posible [3].

CAPITULO IV: BASE DE DATOS MULTIDIMENSIONALES.

4.1. Base de datos Multidimensional.

Una base de datos multidimensional es aquella base de datos en donde su información se almacena en una estructura multidimensional, conformada por tablas de dimensiones y tablas de hechos. Se ha determinado conceptualmente tres variantes de modelamiento, puntualizando las consultas de soporte de decisión, estas son:

- a) **Esquema en estrella (Star Scheme):** el esquema en estrella está formado por una tabla de hechos base y de varias tablas de dimensiones relacionadas a esta, a través de las claves de relación. Es el esquema más simple de interpretar y permite optimizar los tiempos de respuesta ante las consultas de los usuarios. Este esquema, siendo el más eficiente, es soportado por la mayoría de las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener [10].

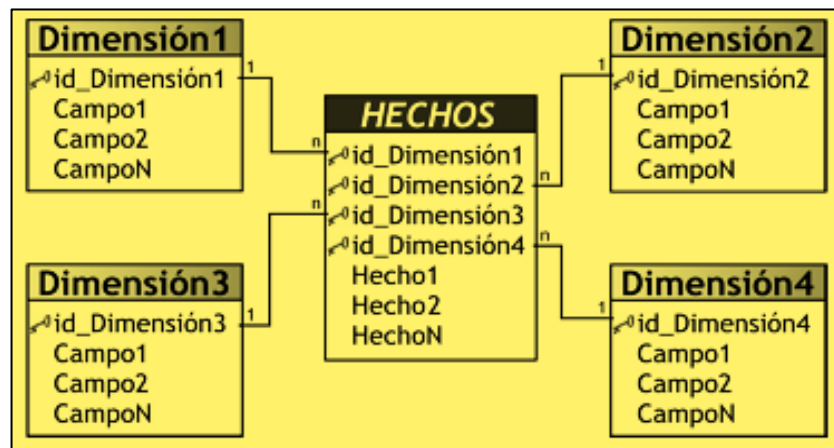


Figura 7 Esquema en Estrella.

- b) **Esquema copo de nieve (Snowflake Scheme):** el esquema copo de nieve consiste en una tabla de hechos central relacionada con una o más tablas de dimensiones, las que a su vez pueden estar relacionadas con otras tablas de dimensiones. Este esquema representa el modelo en estrella pero con las dimensiones organizadas en forma de jerarquías [10].

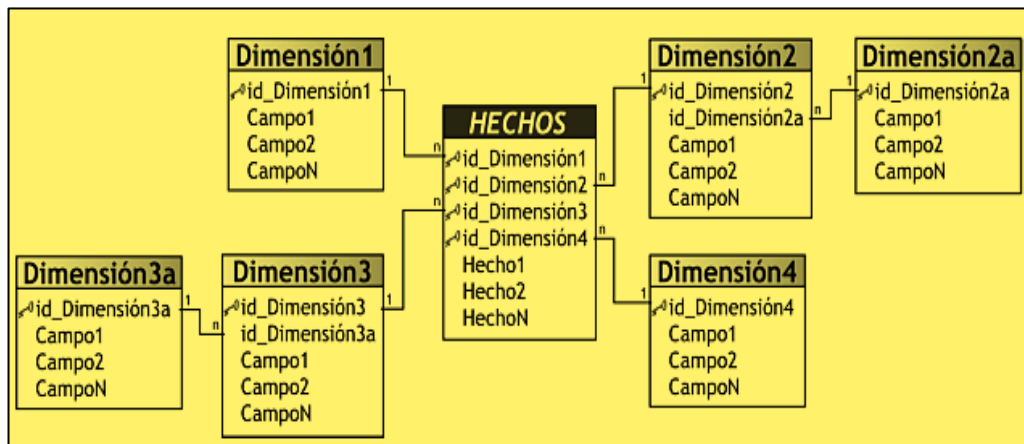


Figura 8 Esquema Copo de Nieve.

c) **Esquema constelación o copo de estrellas (Star flake Scheme):** el esquema constelación está compuesto por una serie de esquemas en estrella. Consta de una tabla de hechos principal y una o más tablas de hechos auxiliares, relacionadas con sus respectivas tablas de dimensiones. La implementación de estos esquemas puede ser diversa, pero generalmente requieren que toda la estructura de datos este desnormalizada totalmente o en un pequeño grado de normalización dependiendo del esquema; esto, con el objetivo de mejorar la ejecución de consultas y el análisis [3].

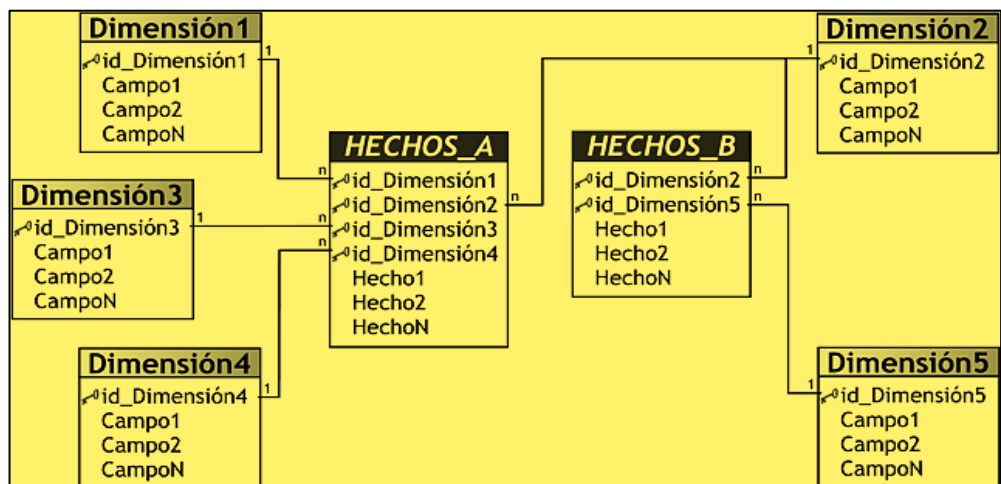


Figura 9 Esquema Constelación o Copo de Nieve.

4.1.1. Tablas de dimensiones.

Las tablas de dimensión permiten definir cómo están los datos organizados lógicamente y describen la forma para analizar el contexto del negocio; además se caracterizan por contener datos cualitativos. Una tabla de dimensión representa los



aspectos de interés mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos, posee un identificador único, denominado clave subrogada y al menos algún campo que describe los criterios de análisis que son relevantes para la organización y por lo general son de tipo texto, es decir estas tablas contienen información del negocio [5].

4.1.1.1. Tabla de dimensión tiempo.

En todo Data Warehouse, la creación y el mantenimiento de una tabla de dimensión tiempo es obligatoria, y la definición de granularidad y estructura de la misma depende de la dinámica del negocio que se está analizando. Es importante tener en cuenta que el tiempo no es solo una secuencia cronológica representada de forma numérica, sino que posee fechas especiales que inciden notablemente en las actividades de la organización. Esto se debe a que los usuarios podrán por ejemplo analizar el número de estudiantes teniendo en cuenta la oferta académica, carrera, paralelo, etc.

4.1.1.2. Dimensiones lentamente cambiantes con el tiempo.

Las dimensiones lentamente cambiantes o SCD (Slowly Changing Dimensions) son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, o implique a un solo registro o la tabla completa. Cuando ocurren estos cambios, se puede optar por seguir alguna de estas dos grandes opciones: registrar el historial de cambios o reemplazar los valores que sean necesarios.

Inicialmente Ralph Kimball planteó tres estrategias a seguir cuando se tratan las SCD: tipo 1, tipo 2 y tipo 3; pero a través de los años la comunidad de personas que se encargaba de modelar bases de datos profundizó las definiciones iniciales e incluyó varios tipos SCD más, por ejemplo: tipo 4 y tipo 6 [3]. A continuación se detallan la SCD tipo 1, y tipo 2 que son las más utilizadas.

- **SCD Tipo 1: Sobrecribir:** Este tipo es el más básico y sencillo de implementar, ya que si bien no guarda los cambios históricos, tampoco requiere ningún modelado especial y no necesita que se añadan nuevos registros a la tabla. En este caso cuando un registro presente un cambio en alguno de los valores de sus campos, se debe proceder simplemente a actualizar el dato en cuestión sobre escribiendo el antiguo [3].
- **SCD Tipo 2: Añadir fila:** esta estrategia requiere que se agreguen algunas columnas adicionales a la tabla de dimensión, para que almacenen el historial de cambios [3]. Las columnas que suelen agregarse son:

- **FechaInicio:** fecha desde que entró en vigencia el registro actual. Por defecto suele utilizarse una fecha muy antigua, ejemplo: "01/01/1000".
- **FechaFin:** fecha en la cual el registro actual dejó de estar en vigencia. Por defecto suele utilizarse una fecha muy futurista, ejemplo: "01/01/9999".
- **Versión:** número secuencial que se incrementa cada nuevo cambio. Por defecto suele comenzar en "1".

4.1.2. Tablas de hechos.

Las tablas de Hechos son la base central de los esquemas multidimensionales porque contienen datos cuantitativos, los cuales son instantáneos en el tiempo; y permiten al usuario filtrarlos, agruparlos y explorarlos a través de condiciones definidas en las tablas de dimensiones. Toda la información presente en las tablas de hechos constituye el volumen de la bodega y pueden estar compuestos por millones de registros dependiendo del nivel de detalle y la historia que el negocio requiere alcanzar.

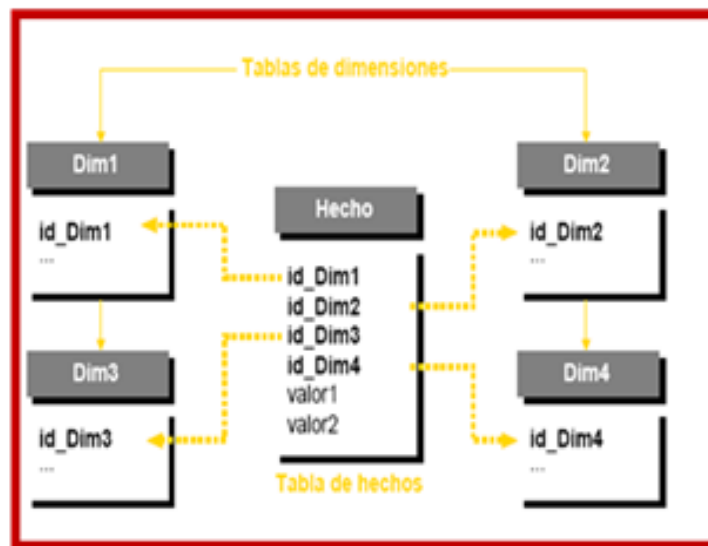


Figura 10 Tabla de hechos.

Para el registro del hecho, este debe poseer una clave primaria única compuesta por las claves primarias de las tablas de dimensiones relacionadas [5]. Los hechos son aquellos datos que residen en una tabla de hechos y que son utilizados para crear indicadores, a través de funciones preestablecidas como:



- **Máximo:** max ()
- **Mínimo:** min ()
- **Conteo:** count ()
- **Suma:** sum ()
- **Promedio:** avg ().

4.1.2.1. Indicadores.

Los indicadores son sumalizaciones efectuadas sobre algún hecho, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el Data Warehouse. El valor que estos adopten estará condicionado por los atributos/jerarquías que se utilicen para analizarlos.

4.1.2.2. Atributos.

Los atributos constituyen los criterios de análisis que se utilizarán para analizar los indicadores dentro de un cubo multidimensional. Los mismos se basan, en su gran mayoría en los campos o datos de referencia de las tablas de dimensiones.

4.1.2.3. Jerarquías.

Una jerarquía representa una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación **padre e hijo (1 - n)**. La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

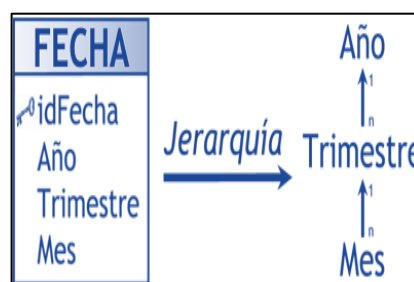


Figura 11 Jerarquía dimensión tiempo.

CAPITULO V: HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS.

5.1. Herramientas de inteligencia de negocios en el mercado.

Según el cuadrante mágico de Gartner Feb 2014, se dispone de las siguientes herramientas orientadas a la inteligencia de negocios en el mercado:

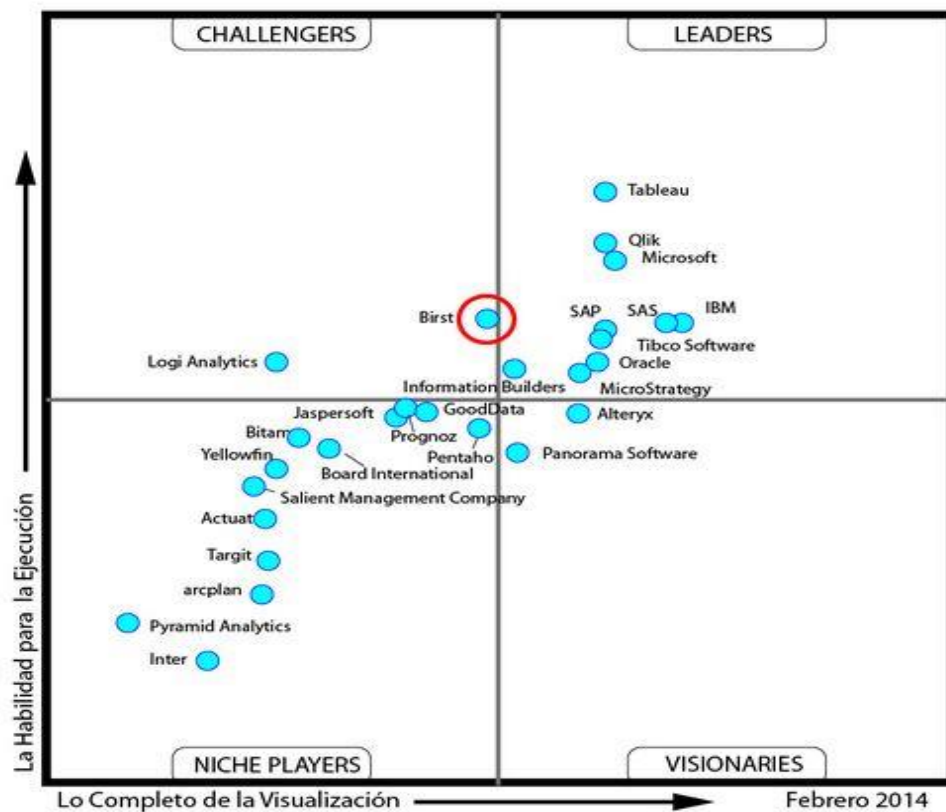


Figura 12 Cuadrante Mágico de Gartner de Plataformas Business Intelligence.

Gartner Inc [12] empresa consultora y de investigación de las tecnologías de la información, define la inteligencia de negocios (BI) y las plataformas analíticas como un software plataforma que entrega 15 capacidades a través de 3 categorías:

1. Entrega de Información.

- Presentación de informes.
- Dashboards.
- Ad hoc informe / consulta.
- Integración con Microsoft office.
- Mobile BI.



2. Integración.

- Gestión de metadatos.
- Herramientas de desarrollo.
- Colaboración (Chat).
- Apoyo a las fuentes grandes de datos.

3. Análisis.

- Visualización interactiva.
- Búsqueda avanzada en el descubrimiento de datos.
- Inteligencia Geoespacial y ubicación.
- Análisis avanzado.
- Procesamiento analítico en línea (OLAP).
- Análisis embebido avanzado [12].

Como se puede apreciar en la Fig. 12, el cuadrante mágico de Gartner está dividido en cuatro secciones, en los que se muestran las capacidades y situación de los desarrolladores de plataformas de inteligencia de negocios, en la parte superior derecha del cuadrante se puede apreciar a los líderes (satisfacen la demanda del mercado), en la parte inferior derecha los visionarios (requieren de apoyo financiero), en la parte superior izquierda los desafiantes (tienen posibilidades de convertirse en líderes) y en la parte inferior izquierda los que están enfocados a un nicho o sección de mercado (capacidad limitada para innovar y superar a otros proveedores).

5.2. Plataformas Open Source enfocadas a Inteligencia de Negocios.

Dentro del conjunto de plataformas orientadas a la inteligencia de negocios mostradas en el cuadrante mágico de Gartner, la mayoría de ellas son corporativas, salvo el caso de Pentaho y Jasper Soft que son dos plataformas que ofrecen la versión empresarial, y la versión libre orientado al ámbito educativo y al segmento de las PYMES. Aunque existen otra gama de productos open source, las más destacadas son Pentaho, Jasper Soft y SpagoBI, ya que ofrecen la suite completa para la implementación de inteligencia de negocios, otras como Talend Open Studio [18] (solo para procesos ETL), y Jedox [17] (solo para reporte), también se las puede considerar en un proyecto BI, a continuación se muestra una comparativa de las herramientas open source más sobresalientes:



TABLA II PLATAFORMAS OPEN SOURCE ENFOCADAS A INTELIGENCIA DE NEGOCIOS.

Módulos	JasperSoft	Pentaho	SpagoBI
Aplicación Server	JBoss	JBoss	JBoss
Autenticación y perfiles de usuario	Acegi	Acegi	Integrated in eXo Portal
Dashboard	JFreeChart	JFreeChart	Openlaszlo
Data Mining	-	Weka	Weka
DBMS	MySQL, Oracle, SQL Server, PostgreSQL, etc	MySQL, Oracle, SQL Server, PostgreSQL, etc	MySQL, Oracle, SQL Server, PostgreSQL, etc
ETL	JasperETL	Pentaho Data Integration	Talend open Studio
Geo Referencia	Google Maps	Google Maps	GEO
Job Scheduler	Quartz	Quartz	Quartz
OLAP	Mondrian&Jpivot	Mondrian&Jpivot	Mondrian&Jpivot
Query by Example	-	-	Hibernate
Reporting	JasperReport	Pentaho Report Designer, JasperReport, BIRT	JasperReport, BIRT
Web Server	Tomcat	Tomcat	Tomcat

CAPITULO XI: OLAP (On Line Analytical Processing).

6.1. Sistemas OLAP (On Line Analytical Processing).

Se puede considerar los sistemas OLAP, como pertenecientes a los sistemas de información para ejecutivos (EIS), utilizados para proporcionar al nivel estratégico información útil para la toma de decisiones [31].

OLAP, es un método que nos permite buscar los datos de diferentes maneras. Con OLAP los datos son clasificados en diferentes dimensiones, las que pueden ser vistas unas con otras en cualquier combinación, para obtener diferentes análisis de los datos que contienen [31].

En un modelo de datos OLAP, la información es vista como cubos, los cuales consisten de categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo de datos multidimensional simplifica a los usuarios la formulación de consultas complejas, arreglar datos en un reporte, cambiar de datos resumidos a datos detallados, filtrar y rebanar los datos en subconjuntos significativos. Dado que las celdas de datos pueden ser fácilmente representadas en un cubo, se pueden tomar rebanadas del mismo para responder a una diversidad de preguntas [31], [32].

6.2. MOLAP (OLAP Multidimensional).

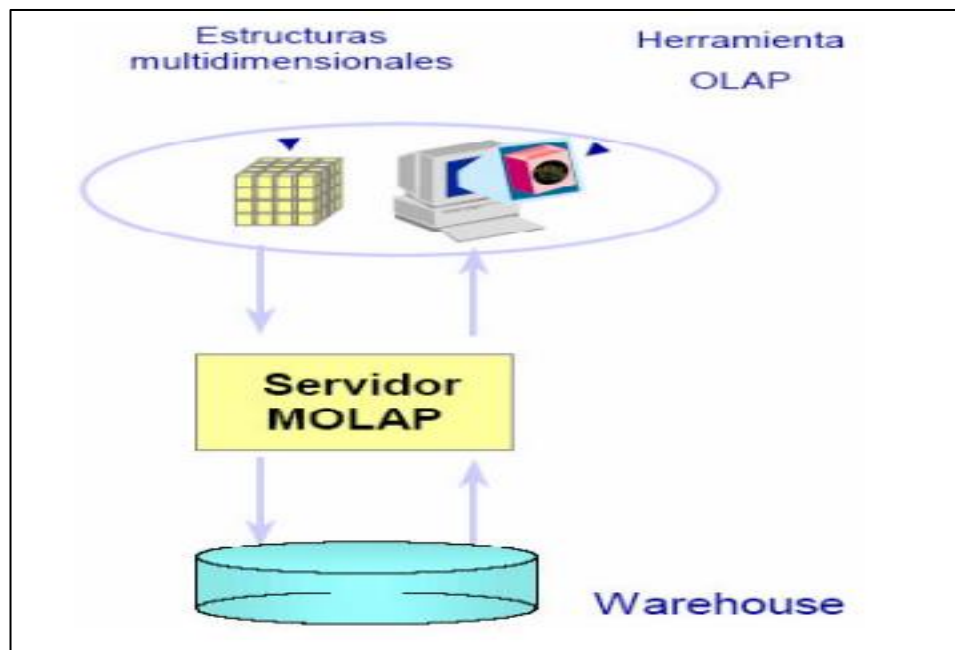


Figura 13 Funcionamiento de los modelos MOLAP.



En estos sistemas los datos se encuentran almacenados en una estructura multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores pre calculados o agregados son la base de las ganancias de desempeño de este sistema [32].

Un sistema MOLAP usa una base datos multidimensionales (BDMD), en la que la información se almacenara dimensionalmente. Este sistema utiliza una arquitectura de dos niveles: la BDMD y el motor analítico [32], [31].

La BDMD es la encargada del manejo, acceso y obtención de los datos, el nivel de aplicación es el responsable de la ejecución de las consultas OLAP.

El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP [31]. A continuación en la tabla III se muestra las ventajas y desventajas que existen al utilizar este tipo de análisis:

TABLA III VENTAJAS Y DESVENTAJAS DE MOLAP.

Ventajas	Desventajas
<ul style="list-style-type: none">- Ofrece buen rendimiento cuando se trabaja sobre datos agregados, totales, subtotales, series temporales y diversos grados de detalle de los datos.- Facilita el estudio a un alto nivel de los datos, al ofrecer una mayor flexibilidad y rapidez de acceso para el análisis de los mismos.- Almacenamiento de datos y consultas bastante eficientes.- La complejidad de la base de datos se oculta al usuario [32].	<ul style="list-style-type: none">- La construcción y poblado de las estructuras multidimensionales puede demandar mucho tiempo.- Están limitados a tener diez o menos dimensiones debido a la complejidad para el manejo de las mismas.- Debe trabajar con volúmenes de datos limitados, menos de 5 GB.- Existen pocas herramientas gratuitas que lo soporten [32].



E. MATERIALES Y MÉTODOS.

Para el desarrollo del presente proyecto de titulación, se recurrió a diferentes técnicas de recolección de información, métodos científicos y procedimientos que la investigación científica ofrece y que son de mucha utilidad. Dentro de los métodos de investigación científica que se manejaron tenemos:

- **Método Analítico.**

Este método nos ayudó a observar y determinar la naturaleza del objeto de estudio, para así poder determinar las causas y efectos del mismo. Además este método nos permitió explicar y comprender un poco más a fondo sobre el funcionamiento de la inteligencia de negocios.

- **Método Inductivo.**

Este método se lo utilizó para estructurar la problemática y justificación, buscando información relacionada con el tema y permitiéndonos dar solución a los diferentes problemas que se presentan en las organizaciones, cuando no poseen herramientas de inteligencia de negocios.

- **Método Deductivo.**

Este método nos permitió obtener información primaria para el desarrollo del proyecto relacionada con inteligencia de negocios. De igual manera se lo utilizó al momento de elegir el tema para el proyecto y el campo de investigación en el que será aplicado. Así mismo será aprovechado para esclarecer conceptos y definiciones que serán de gran importancia durante el transcurso de la investigación.

- **Investigación Bibliográfica.**

La investigación bibliográfica nos sirvió como base para el desarrollo del proyecto, ya que nos proporcionó la fundamentación teórica que nos ayudará a determinar las especificaciones técnicas para el diseño y la implementación de inteligencia de negocios, elaboración de informes, así como también comprender el funcionamiento de la herramienta Pentaho Business Intelligence.

Técnicas de recolección de información.

- **Técnica de entrevista.**

La entrevista se la utilizó para obtener mayor información acerca de las necesidades de información que existen dentro del AEIRNNR, así como las limitaciones que tienen los sistemas de información que se manejan dentro de la misma.



- **Observación directa.**

Esta técnica se la utilizó para tener una observación de campo, para constatar los procesos, los sistemas transaccionales, las fuentes de datos y la problemática que existe dentro del AEIRNNR, en cuanto a la gestión de información académica se refiere.

- **Metodología para inteligencia de negocios.**

Para el desarrollo del proyecto, se utilizó la metodología de inteligencia de negocios HEFESTO V 2.0, la misma que se organiza en cuatro etapas que se describen a continuación:

- 1. Análisis de requerimientos.**

Lo primero que se realizó fue identificar los requerimientos del usuario a través de preguntas que expliciten los objetivos de la organización. Luego, se analizaron estas preguntas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. Finalmente se confeccionó un modelo conceptual, en donde se podrá visualizar el resultado obtenido en este primer paso.

- 2. Análisis de los OLTP.**

Durante esta fase se analizaron las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos.

- 3. Modelo lógico del DW.**

En esta fase se confeccionó el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se definió el tipo de modelo que se utilizará y luego se llevaron a cabo las acciones propias al caso para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizaron las uniones pertinentes entre estas tablas.

- 4. Integración de datos.**

Una vez construido el modelo lógico del DW, se procedió a poblarlo con datos, para ello se utilizó técnicas de limpieza y calidad de datos, procesos ETL, etc.; luego se definieron las reglas y políticas para su respectiva actualización, así como también los procesos que la llevarán a cabo. Una vez poblado de datos el DW, se presentó la información contenida en el mismo mediante reportes y cubos OLAP.



F. RESULTADOS.

Para el desarrollo del presente proyecto se definieron tres fases, las mismas que nos ayudaron a culminarlo con éxito. A continuación se detallan los resultados obtenidos en cada una de estas fases, con el fin de verificar el cumplimiento de las mismas.

1. Recopilación de Información.

1.1. Analizar y comprender la Naturaleza del Negocio.

Para comprender de mejor manera los procesos del AEIRNNR, se realizó un análisis e investigación de su entorno, aplicando técnicas de recolección de información como la observación directa y entrevista realizada al director de la misma (véase Anexo 2).

En la actualidad, el AEIRNNR es administrada por el Ing. Jorge Gahona Pacheco, esta área de la Universidad Nacional de Loja, alberga un promedio de 400 a 700 estudiantes por oferta académica. Para cumplir con la misión, visión en bien de los estudiantes se encuentra integrada por diferentes carreras, las cuales contribuyen directamente con los procesos que agregan valor, para tener una visión más clara de la estructura del AEIRNNR, en el Anexo 9 se nos muestra la estructura académica de la misma.

Dentro de los diferentes departamentos que conforman el área, específicamente el departamento de dirección, no dispone de herramientas automatizadas que ayuden a profundizar el nivel de conocimiento de la información académica, lo que produce un incompleto e ineficiente soporte de respaldo, para apoyar el proceso de toma de decisiones administrativas. Otros problemas significativos que afectan a la dirección de esta área de la Universidad Nacional de Loja son los siguientes:

- Pérdidas de tiempo, tanto para generar consultas de gran tamaño como para la entrega de respuestas.
- Carencia de una plataforma en la que se pudiera interactuar con la información académica histórica.
- Carencia de herramientas que otorguen reportes en lapsos breves de tiempo.

1.2. Obtener los Requerimientos e Indicadores claves de rendimiento.

El AEIRNNR trabaja con un indicador clave de rendimiento como los alumnos matriculados por oferta académica y carrera, donde el proceso se refiere al año de admisión en que los alumnos postularon y se matricularon. Esta medida es la base para la entrega de diversas métricas importantes, tanto dentro del área como también



para la universidad. Los datos que son derivados en base a este indicador pueden ir detallándose más y más (dependiendo de la necesidad de información solicitada). Por ejemplo, el valor cantidad de docentes titulares o contratados pertenecientes a una determinada oferta académica, así como la cantidad de alumnos aprobados o reprobados por oferta académica y carrera, distribución de alumnos por sexo, etc. A continuación se detallan las necesidades o requerimientos de información que existen dentro del área.

TABLA IV IDENTIFICACIÓN DE REQUERIMIENTOS.

Entrada	Descripción
Tema.	Área de la Energía, las Industrias y los Recursos Naturales no Renovables (AEIRNNR).
Audiencia.	Ing. Jorge Gahona Pacheco Director del Área de la Energía, las Industrias y los Recursos Naturales No Renovables.
Propietario.	Universidad Nacional de Loja.
Necesidad del usuario	Datamarts (Docentes-Estudiantes), Cubos OLAP, Reportes a demanda con la información académica del AEIRNNR.
Preguntas a responder (Requerimientos).	<ul style="list-style-type: none">• Total de estudiantes matriculados en una oferta académica.• Distribución de estudiantes por género, en una oferta académica y carrera.• Alumnos distribuidos en cada carrera del total de los alumnos matriculados en una oferta académica.• Promedios de notas por carrera en una determinada oferta académica.• Tasa de retención estudiantil.• Porcentaje de aprobados y reprobados en una oferta académica y carrera.• Total de docentes en una determinada oferta y su distribución por carreras• Tasa de profesores clasificados por categoría en una determinada carrera.• Nivel Titularidad de profesores• Formación de posgrado
Beneficios del Negocio	<ul style="list-style-type: none">• Agilización al generar reportes.• Información útil y relevante• Ahorro de tiempo
Mecanismos de entrega.	<ul style="list-style-type: none">• Reportes en formato PDF o Excel.
Fuentes de información.	<ul style="list-style-type: none">• Ing. Jorge Gahona Pacheco: Director del AEIRNNR.



1.3. Analizar las distintas fuentes de datos del negocio.

La información académica de la Universidad Nacional de Loja, es procesada por el Sistema de Gestión Académica (SGA), y este a su vez almacena la información dentro de bases de datos relacionales, que son administradas por la Unidad de Telecomunicaciones e Información (UTI) perteneciente a la misma universidad.

La gran mayoría de los datos internos académicos del AEIRNR, se encuentran alojados dentro de estas bases de datos relacionales, es decir aquí se contempla la información de los alumnos y docentes, así como los datos personales, el rendimiento académico de los alumnos de todas y cada una de las carreras pertenecientes al AEIRNR, entre otros aspectos relacionados con la parte académica de la misma. Además uno de los servicios que proporciona la UTI, es el SGAServices (Disponible en: <http://ws.unl.edu.ec>), el mismo que ha sido desarrollado como proyecto de tesis, con el fin de proporcionar información académica contenida dentro del SGA, para que pueda ser aprovechada por aplicaciones que tengan fines académicos, razón por la cual se lo ha elegido a este último como principal OLTP para el desarrollo del presente proyecto, ya que por políticas de seguridad de la misma universidad, no se puede dar acceso a las bases de datos del SGA.

Se debe mencionar que el SGAServices, contiene información a partir del año 2008 en adelante y que existen datos de carreras que ya no se encuentran ofertadas dentro del área, y en muchos casos esta información no es completa, por tal motivo se consideró únicamente las cuatro carreras principales que aún se ofertan en el área (Ingeniería en Sistemas, Electrónica y Telecomunicaciones, Geología y Ordenamiento Territorial, Electromecánica).

2. Selección de la Metodología.

Dentro del ámbito de inteligencia de negocios, existen gran variedad de metodologías enfocadas a este campo, pero las más sobresalientes son las que proponen los padres del Data Warehousing Bill Inmon y Ralph Kimball, aunque existe la metodología HEFESTO v2.0, que es un tipo de metodología híbrida, ya que permite integrar los conceptos de Inmon como de Kimball. A continuación se muestra una comparativa de dichas metodologías:



TABLA V METODOLOGÍAS PARA INTELIGENCIA DE NEGOCIOS.

Metodología	Enfoque	Características	Resumen
HEFESTO 2.0	DW	Construcción de DW	<ol style="list-style-type: none">1. Análisis de requerimientos.2. Análisis OLTP.3. Modelo lógico DW.4. Integración de datos
KIMBALL	DW	Diseño de DW	<ol style="list-style-type: none">1. Planeación del proyecto.2. Requerimientos del negocio.3. Diseño Arquitectura Técnica.4. Selección e instalación de herramientas.5. Modelado dimensional.6. Diseño físico.7. Especificaciones.8. Desarrollo herramienta analítica.9. Área de Staging.10. Desarrollo.11. Mantenimiento.
INMON	DW	Diseño de DW	<ol style="list-style-type: none">1. Implementación DW.2. Integración de Datos.3. Pruebas.4. Programación sobre datos.5. Diseño DSS.6. Análisis resultados.7. Entendimiento



La metodología a considerar para el desarrollo del proyecto es la metodología HEFESTO, entre las características más sobresalientes para elegir esta metodología podemos mencionar:

- Se aplica tanto para almacén de datos como para Data Marts.
- Independencia de la estructura física que contengan el almacén de datos.
- Independencia de la plataforma que se utilice para la implementación del Data Mart.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Específica puntualmente los pasos a seguir en cada fase.

3. Desarrollo del Data Warehouse aplicando la metodología HEFESTO.

3.1. Modelo lógico y físico del Data Warehouse.

Para el desarrollo del DW del AEIRNNR, se consideró la arquitectura MD (Multi Dimensional) de Ralph Kimball, debido a que esta arquitectura nos permite construir uno o más Data Marts enfocadas en unidades del negocio individuales (Estudiantes - Docentes), y estos a su vez conforman el DW de la organización. Otra de las características adicionales de esta arquitectura, es la utilización del modelamiento de datos dimensional que se puede llevar a cabo en cada Data Mart, y el significativo ahorro de tiempo y costos en su implantación. Para este caso específico, se construyó un Data Mart encaminado a la parte de estudiantes, y un segundo Data Mart para la parte de docentes, todo este proceso se lo realizó basándonos en la metodología HEFESTO que ha sido definida anteriormente.

Se debe destacar que el desarrollo de un DW es iterativo, la flexibilidad para responder a cambios de requerimientos del negocio es importante, al entrar el DW en un entorno de producción. Si se lo desarrollaría de forma secuencial y en cascada, se anularía la característica de flexibilidad, ya que se lo limitaría al DW a los requerimientos iniciales que han sido definidos al comienzo del desarrollo. Se debe justificar que los requerimientos se definirán una única vez, y serán utilizados con el propósito de explicar cómo usar la metodología de desarrollo para construir el DW y dar solución al problema planteado.

Otro aspecto importante dentro de un proyecto de inteligencia de negocios, es la elección de la plataforma de desarrollo, en este caso se ha seleccionado la plataforma



Pentaho Business Intelligence, la misma que dispone de todas las herramientas necesarias para adoptar una arquitectura de inteligencia de negocios (Arquitectura Multidimensional MD), sirviendo de apoyo desde los procesos ETL, hasta la entrega de información al usuario final a través de reportes, cubos OLAP, etc. Entre las características más relevantes de Pentaho tenemos:

- Open Source (no se pagará ningún tipo de licencia).
- Dispone de información en línea.
- Provee todos los módulos necesarios para poder realizar todo el proceso Data Warehousing (ETL, OLAP, Reporteo a Demanda), es decir, dispone de una cobertura total de todas las necesidades de inteligencia de negocios.
- La conexión con Web Services, es una característica esencial de Pentaho para el desarrollo del presente proyecto, ya que se ha tomado como principal OLTP al SGASWebServices de la Universidad Nacional de Loja.
- Facilidad para poder trabajar con diversos motores de bases de datos (PostgreSQL, MySQL, SQLite, Oracle, etc.), así como también archivos planos (Excel, XML, texto, etc.).
- Soporte y capacitación en línea.
- Está basado en estándares.
- Es 100 % JAVA con amplio soporte multiplataforma.
- Basado en repositorio, permite reusar componentes de transformación.

El módulo Pentaho Data Integration, se lo utilizó para los procesos de Extracción, Transformación y Carga (ETL), el módulo Pentaho Report Designer para generar reportes, y para generar cubos OLAP se utilizó el módulo Pentaho Schema Workbench, ambos nos permitieron generar reportes avanzados mediante la conexión a diferentes tipos de bases de datos.

Finalmente, para la entrega de información al usuario final, se utilizó el módulo denominado Pentaho BIService, este módulo constituye el centro de entrega de información, permitiendo mantener los reportes de análisis en una estructura de directorios creada previamente, así como también los usuarios con sus distintos roles y permisos asignados. A continuación se muestra el modelo de componentes de inteligencia de negocios, definido en base a la plataforma Pentaho Business Intelligence.

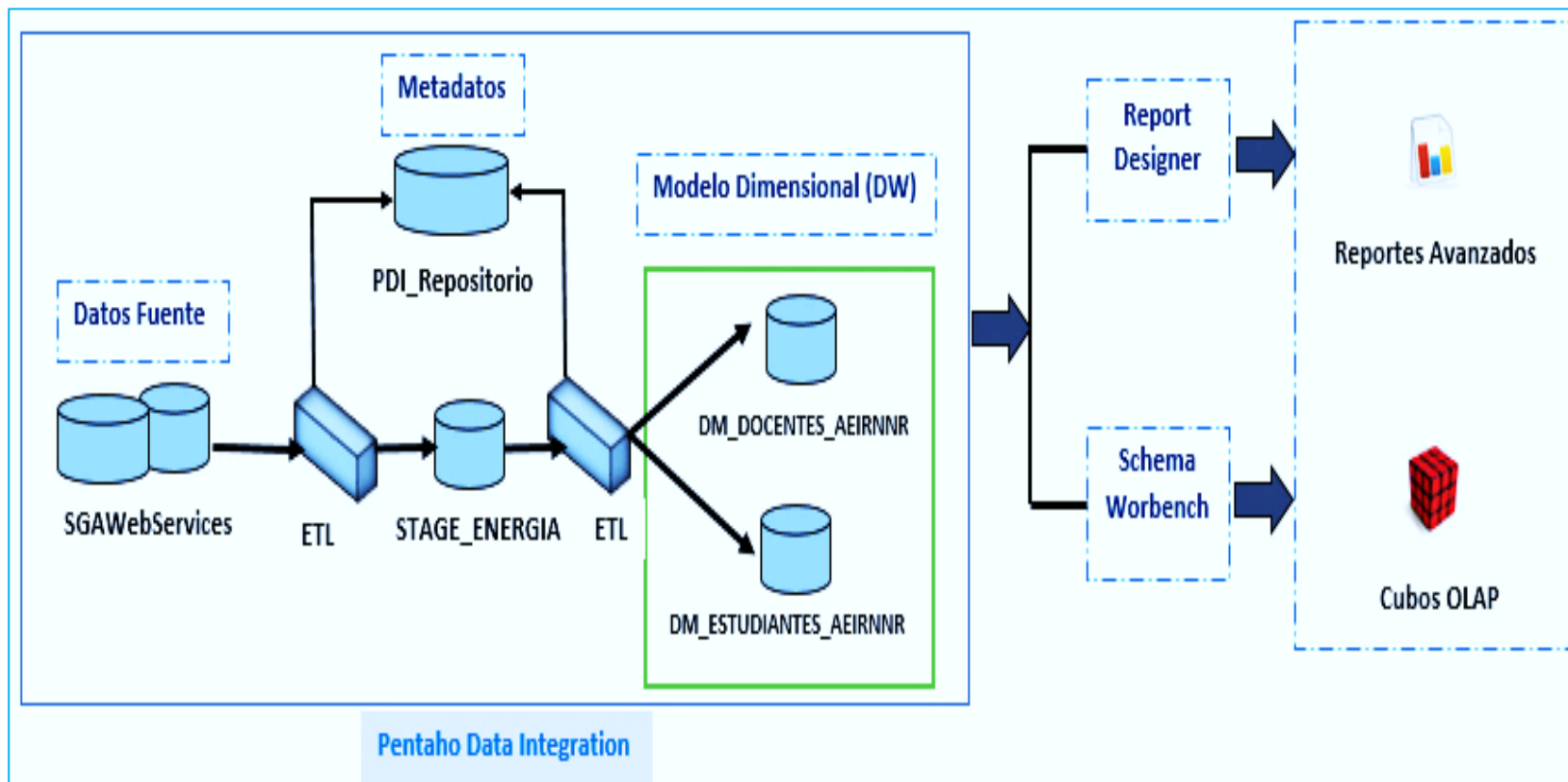


Figura 14 Modelo de Componentes de Inteligencia de Negocios para el AEIRNNR.



3.1.1. Etapa 1: Análisis de Requerimientos.

3.1.1.1. Identificar indicadores y perspectivas de análisis.

Una vez obtenidas las necesidades de información (Véase *tabla IX*), se analizaron cada una de ellas y se consideró únicamente aquellas que estén contenidas dentro del principal OLTP que se dispone para la realización del presente proyecto (SGAWebServices), a continuación se detalla cuáles son sus respectivos indicadores y perspectivas.

TABLA VI INDICADORES Y PERSPECTIVAS DE ANÁLISIS.

Preguntas	Indicadores	Perspectivas
Total de estudiantes matriculados en una oferta académica.	Tasa estudiantes.	Oferta académica.
Distribución de estudiantes por género, en una oferta académica y carrera	Distribución por género.	Oferta académica, carrera, género.
Alumnos distribuidos en cada carrera del total de los alumnos matriculados en una oferta académica	Distribución alumnos.	Oferta académica, Carrera.
Promedios de notas por carrera en una determinada oferta académica.	Promedios	Oferta académica, Carrera.
Tasa de retención estudiantil.	Tasa de retención	Oferta académica, Carrera.
Tasa de aprobados y reprobados en una oferta académica, carrera.	Tasa Aprobados / reprobados.	Oferta académica, Carrera.
Total de docentes por oferta académica y carrera	Distributivo docente.	Oferta académica, Carrera.
Tasa de profesores clasificados por categoría en una determinada Oferta Académica y Carrera.	Tasa categoría docente.	Profesor, Carrera, Categoría, Oferta académica.
Formación de postgrado de los docentes de una carrera	Formación de postgrado.	Profesor, Título.

3.1.1.2. Modelo conceptual.

Los indicadores y perspectivas definidos anteriormente, se los agrupo por docentes y alumnos respectivamente, y en base a esto se realizó dos modelos conceptuales, los mismos que nos permiten tener un alto nivel de definición de los datos y comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para realizar el análisis y cuál es la relación que existe entre ellos.

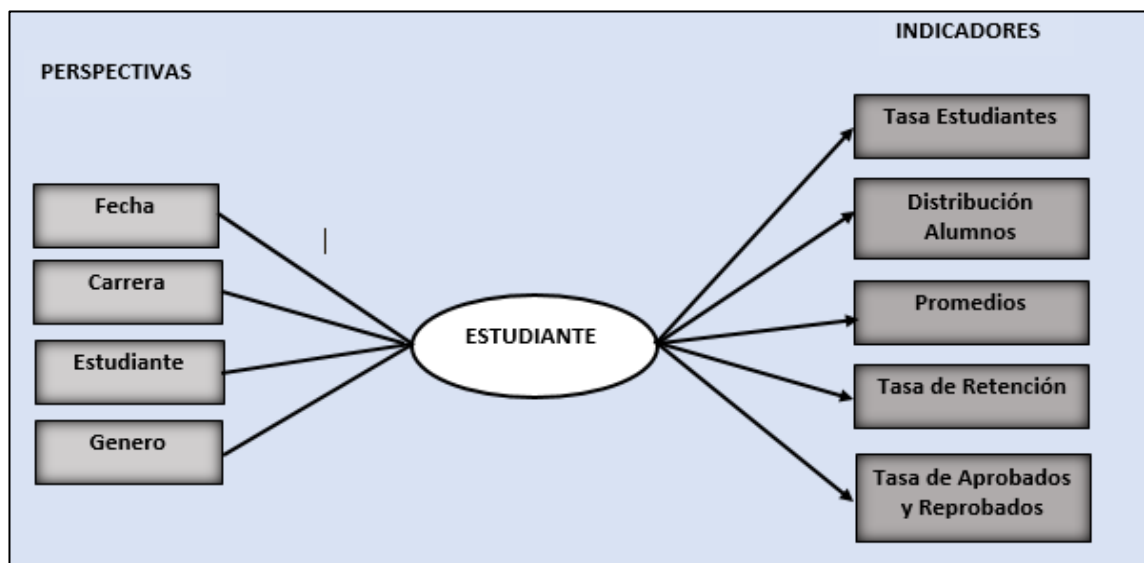


Figura 15 Modelo conceptual DM_ESTUDIANTES_AEIRNNR.

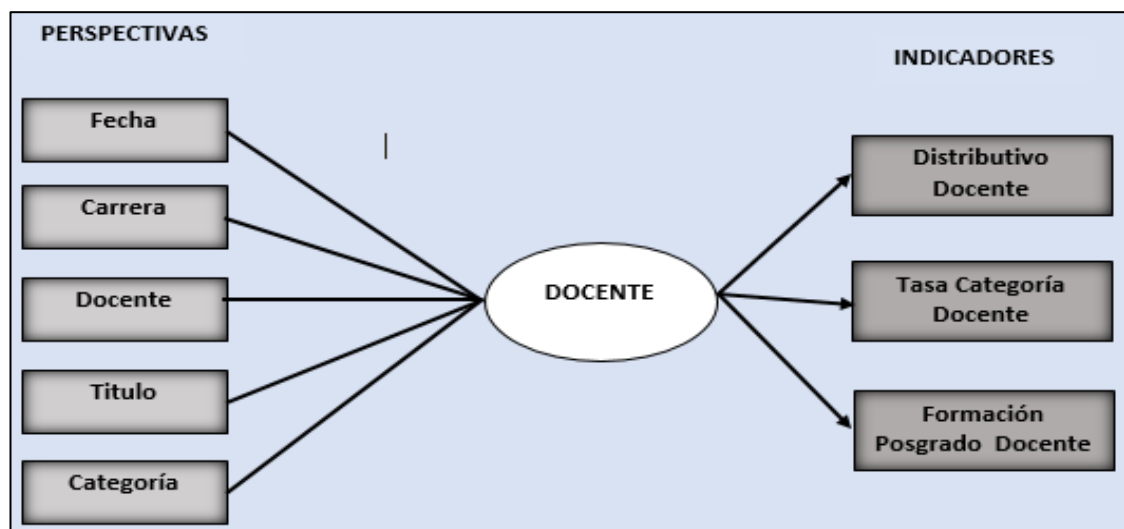


Figura 16 Modelo Conceptual DM_DOCENTES_AEIRNNR.



3.1.2. Etapa 2: Análisis de los OLTP.

El OLTP considerado para la realización del proyecto es el SGAWebServices, el mismo que dispone de una biblioteca de métodos, servicios y distintas funciones que retornan información académica. La información que devuelve cada uno de los métodos del SGAWebServices se encuentra en formato JSON, por tal motivo se deberá realizar cambios de formato de información para poder ser llevada a las fuentes finales, otro de los aspectos a considerar del SGAWebServices es la falta de actualización de la información, así como también la inconsistencia en los datos que devuelven muchos de los métodos. A continuación se mostrará cada una de las categorías de métodos pertenecientes al SGAWebServices:

TABLA VII MÉTODOS SGAWEBSERVICES INFORMACIÓN INSTITUCIONAL.

Método	Datos que retorna
sgaws_lista_areas	Retorna una lista de las áreas que conforman la universidad
sgaws_datos_area	Retorna los datos de las siglas del área ingresada
sgaws_carreras_area	Retorna las carreras que forman parte de una área
sgaws_datos_carreras	Retorna una lista de carreras con sus datos en una oferta académica
sgaws_modulos_carrera	Retorna los datos de todos los módulos de una carrera en una oferta académica
sgaws_paralelos_carrera	Retorna todos los paralelos de una carrera en una oferta académica

TABLA VIII MÉTODOS SGAWEBSERVICES INFORMACIÓN PERSONAL.

Método	Datos que retorna
sgaws_datos_docente	Retorna los datos o información del docente, que corresponden a la cédula ingresada
sgaws_datos_estudiante	Retorna los datos o información del estudiante, que corresponden a la cédula ingresada.
sgaws_datos_usuario	Retorna los datos del usuario ingresado e indica el tipo de usuario estudiante / docente o ambos que corresponden a la cédula ingresada



TABLA IX MÉTODOS SGAWEB SERVICES INFORMACIÓN ACADÉMICA.

Método	Datos que retorna
sgaws_periodos_lectivos	Retorna una lista de períodos lectivos
sgaws_ofertas_academicas	Retorna una lista de ofertas académicas en un periodo lectivo
sgaws_fechas_matriculaoa	Retorna las fechas de matrículas ordinaria, extraordinaria y especial de una oferta académica
sgaws_estadoestudiantes_paralelo	Retorna la lista de estudiantes de un paralelo con el estado de matrícula (aprobada, reprobado, matriculada), de acuerdo a la oferta académica, módulo, carrera y paralelo
sgaws_plan_estudio	Retorna el plan de estudio asignado o un paralelo en una oferta académica
sgaws_carga_horaria_docente	Retorna la carga horaria asignada a un docente en una oferta académica
sgaws_notas_estudiante	Retorna las notas de un estudiante en cada unidad, la carrera, módulo, paralelo al que pertenece y estado de la matrícula en la oferta académica
sgaws_carreras_estudiante	Retorna los datos de un estudiante y las carreras en las que está matriculado
sgaws_reporte_matricula	Retorna los datos de un estudiante, oferta académica, nota y porcentaje de asistencia y estado de la matrícula del estudiante en esa oferta.



3.1.2.1. Construcción de los indicadores.

Después de realizado el análisis del SGAWebServices de la Universidad Nacional de Loja, se procedió a construir los indicadores claves para la gestión académica del AEIRNNR, a continuación se detallan cada uno de ellos.

1. Tasa de estudiantes por Oferta Académica.

- **Hechos:** Estudiante.
- **Función:** Sum (estudiantes ofertas académicas).

2. Distribución por género.

- **Hechos:** Estudiante.
- **Función:** Sum (estudiantes por género).

3. Distribución de alumnos por carrera.

- **Hechos:** Estudiante.
- **Función:** Sum (estudiantes Carreras).

4. Promedios.

- **Hechos:** Notas.
- **Función:** Avg (Calificación).

5. Tasa de retención.

- **Hechos:** Estudiante.
- **Función:**

Tasa de retención = $100 * (\text{Número de estudiantes matriculados en la carrera en la oferta actual y que fueron admitidos dos años antes} / (\text{Número total de estudiantes que fueron admitidos en la carrera dos años antes de la oferta actual}))$.

6. Tasa de aprobados y reprobados.

- **Hechos:** Estado.
- **Función:** Sum (Estudiantes aprobados), Sum (Estudiantes Reprobados).

7. Total docentes por oferta académica y carrera.

- **Hechos:** Docente.
- **Función:** Sum (Docentes por oferta y carrera).



8. Tasa de profesores por categoría.

- **Hechos:** Docente.
- **Función:** Sum (docentes por categoría).

9. Formación de Posgrado.

- **Hechos:** Docente.
- **Función:** Sum (docentes con título de cuarto nivel).

3.1.2.2. Establecer Correspondencias.

Las correspondencias OLTP para el proyecto se derivan del SGAWebServices de la Universidad Nacional de Loja, en las tablas (XII, XIII, IX) descritas anteriormente se nos muestra cada uno de los métodos pertenecientes al mismo.

Las relaciones identificadas entre los métodos del SGAWebServices y las perspectivas e indicadores, fueron las siguientes:

- El método **sgaws_datos_estudiante**, está relacionado con la perspectiva **Estudiante, Genero** y con los indicadores **Tasa Estudiantes, Distribución Alumnos, Tasa de Retención**.
- El método **sgaws_estado_estudiante_paralelo**, está relacionado con el indicador **Tasa de Aprobados y Reprobados**.
- El método **sgaws_ofertas_academicas**, está relacionado con la perspectiva **Fecha**.
- El método **sgaws_periodos_lectivos**, está relacionado con la perspectiva **Fecha**.
- El método **sgaws_carreras_area**, está relacionado con la perspectiva **Carrera**.
- El método **sgaws_notas_estudiante**, está relacionado con el indicador **Promedios**.
- El método **sgaws_datos_docente**, está relacionado con la perspectiva **Docente, Título, Categoría** y con los indicadores, **Distributivo Docente, Tasa Categoría Docente, Formación Posgrado Docente**.

A continuación, en las Fig. 17 – 18, se mostrará las respectivas correspondencias que existen entre los modelos conceptuales definidos anteriormente y los métodos del SGAWebServices.

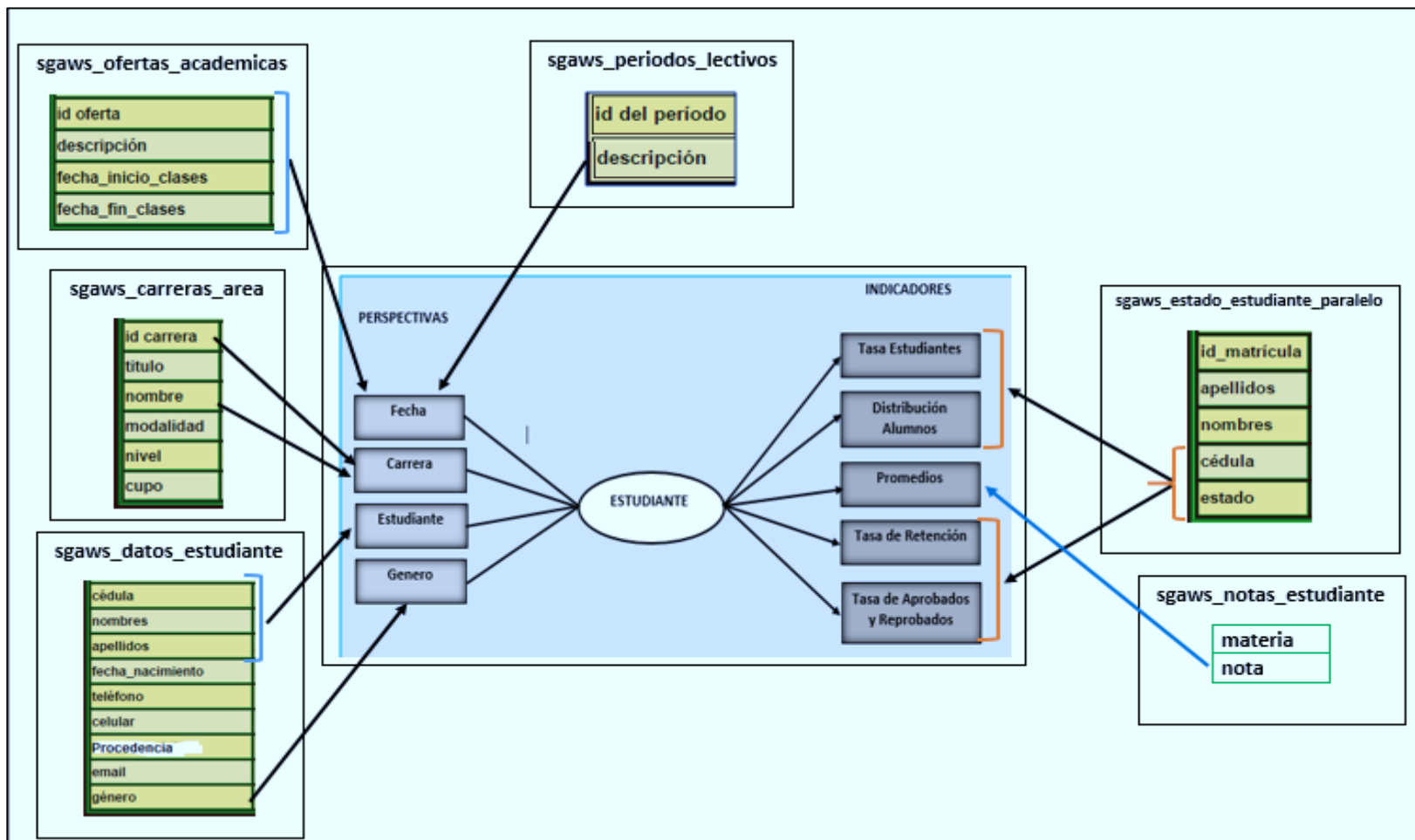


Figura 17 Correspondencias entre métodos del SGASWebServices y el modelo conceptual DM_ESTUDIANTE_AEIRNNR.

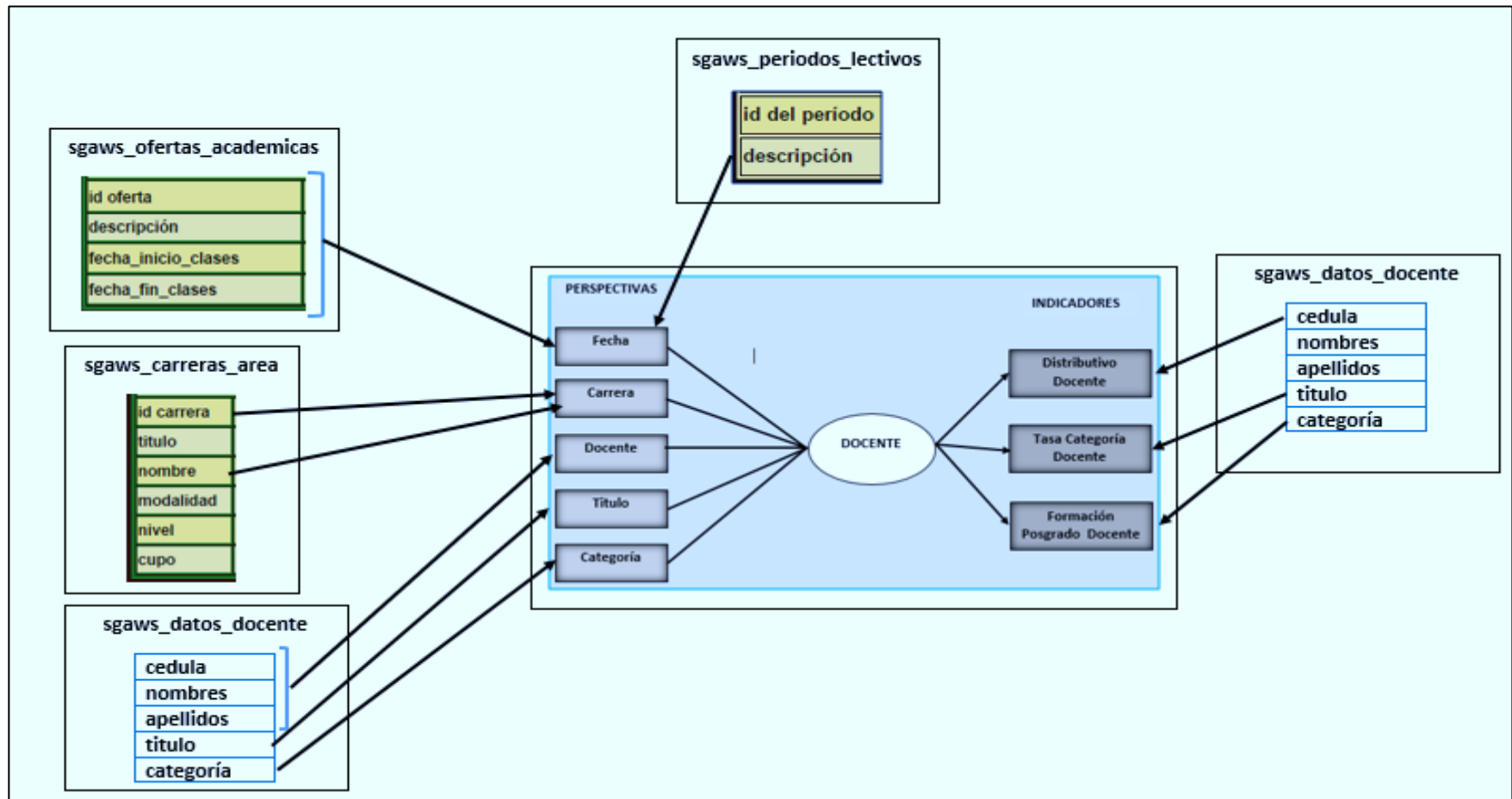


Figura 18 Correspondencias entre métodos del SGAServices y el modelo conceptual DM_DOCENTES_AEIRNNR.



3.1.2.3. Nivel de granularidad.

Una vez establecidas las correspondencias, se analizaron los campos que retornan cada método y cada perspectiva a la que estos hacen referencia, para intuir los significados de cada campo que devuelve.

- Para la perspectiva **ESTUDIANTE Y GENERO** los datos disponibles son los siguientes:

Método: sgaws_datos_estudiante (cedula): este método recibe como parámetro la cedula del estudiante y retorna como salida los siguientes datos:

TABLA X SALIDA MÉTODO SGAWS_DATOS_ESTUDIANTE.

Datos que retorna.	
cedula	Número de cedula que identifica a un estudiante
nombres	Nombres del estudiante del que se ingreso la cedula
apellidos	Apellidos del estudiante del que se ingreso la cedula
fecha nacimiento	Fecha de nacimiento del estudiante
celular	Número de teléfono movil del estudiante
género	Genero del estudiante: masculino o femenino

- Para la perspectiva **CARRERA** los datos disponibles son los siguientes:

Método: sgaws_carreras_area (siglas): este método recibe como parámetro las siglas del área (AEIRNNR) y nos retorna la siguiente información con respecto a las carreras:

TABLA XI SALIDA MÉTODO SGAWS_CARRERAS_AREA.

Datos que retorna.	
id carrera	Número que identifica a una carrera
título	Se refiere al título otorgado por la carrera
nombre	Denominación que tiene una carrera
modalidad	Modo de estudio: presencial, semipresencial o distancia
nivel	Grado de estudio: pregrado, posgrado, tecnico tecnológico
cupó	Numero máximo de estudiantes que una carrera puede tener en primer año



- Para la perspectiva **DOCENTE, TÍTULO, CATEGORÍA** los datos disponibles son los siguientes:

Método: `sgaws_datos_docente (cedula)`: este método recibe como parámetro la cédula del docente y retorna la siguiente información con respecto al docente:

TABLA XII SALIDA MÉTODO SGAWS_DATOS_DOCENTE.

Datos que retorna.	
nombres	Nombres del docente del que se ingreso la cédula
apellidos	Apellidos del docente del que se ingreso la cédula
cédula	Cédula del docente del que se ingreso la cédula
título	Título del docente del que se ingreso la cédula
categoría laboral	Categoría laboral del docente del que se ingreso la cédula: titular o contratado.

- Para la perspectiva **TIEMPO**, que es la que determina la granularidad de los Data Marts, los datos de los que se dispone son los siguientes:
 - Granularidad para esta perspectiva es:
 - Período Lectivo (año).
 - Oferta Académica (Semestre).
 - Descripción de los métodos asociados al SGASWebServices y a la perspectiva tiempo:

Método: `sgaws_periodos_lectivos ()`: este método no recibe parámetros, devuelve la siguiente información relacionada con los periodos lectivos:



TABLA XIII SALIDA MÉTODO SGAWS_PERIODOS_LECTIVOS.

Datos que retorna.	
id periodo	Número que identifica a un periodo lectivo
descripción	Denominación que tiene un período lectivo

Método: sgaws_ofertas_academicas (id_periodo): este método recibe como parámetro el id de un periodo lectivo, y nos devuelve las ofertas académicas pertenecientes al periodo ingresado. Los datos que devuelve son:

TABLA XIV SALIDA MÉTODO SGAWS_OFERTAS_ACADEMICAS.

Datos que retorna.	
id oferta	Número que identifica a un periodo lectivo
descripción	Denominación que tiene una oferta académica
fecha_inicio-clases	Fecha de inicio de clases en una oferta académica
fecha_fin_clases	Fecha de fin de clases en una oferta académica

Después de analizar la información que nos devuelve cada método, se procedió a seleccionar los datos que se considera de interés para analizar los indicadores antes expuestos. A continuación se muestran los resultados obtenidos:

- Perspectiva “ESTUDIANTE”.
 - CEDULA.
- Perspectiva “DOCENTE”.
 - CEDULA
 - NOMBRES
 - APELLIDOS



- Perspectiva “CARRERA”.
 - ID CARRERA.
 - NOMBRE.
- Perspectiva “GENERO”.
 - GENERO.
- Perspectiva “TITULO”.
 - TITULO.
- Perspectiva “CATEGORÍA”.
 - CATEGORÍA.
- Perspectiva “FECHA”.
 - OFERTA ACADEMICA (Descripción).
 - PERIODO LECTIVO (Descripción).
 - FECHA_INICIO.
 - FECHA_FIN.

3.1.2.4. Modelo conceptual ampliado.

El modelo conceptual ampliado, resultado del análisis realizado a cada Data Mart definidos anteriormente, se muestra a continuación:

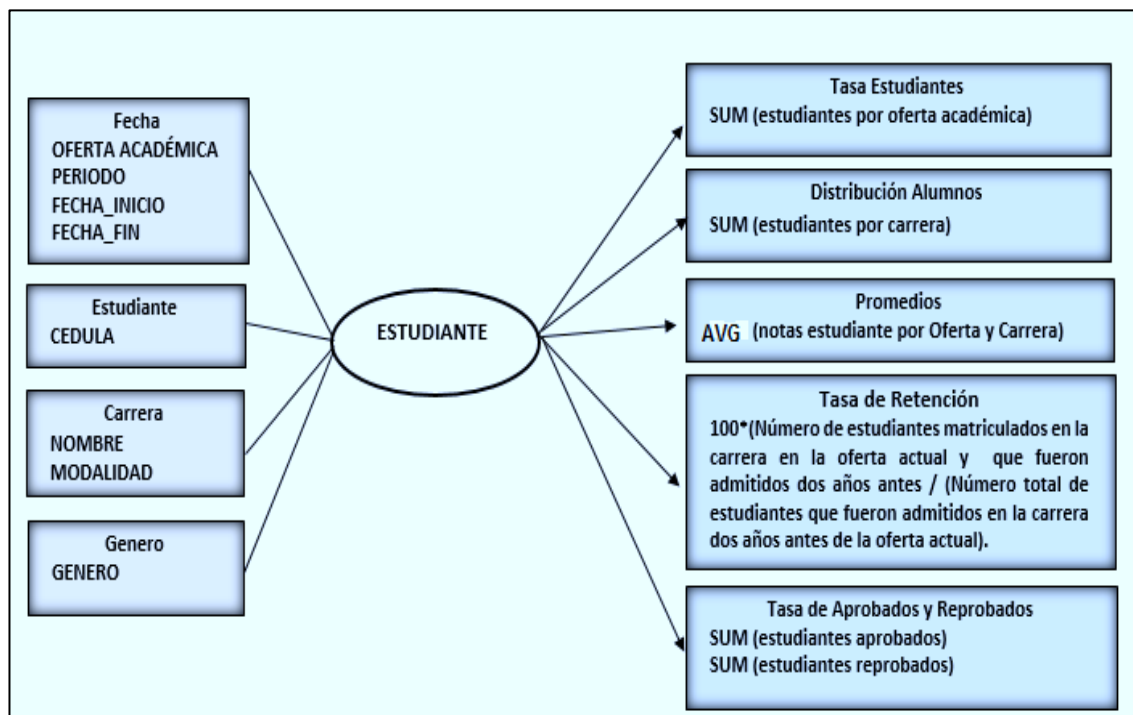


Figura 19 Modelo conceptual ampliado DM_ESTUDIANTES_AEIRNNR.

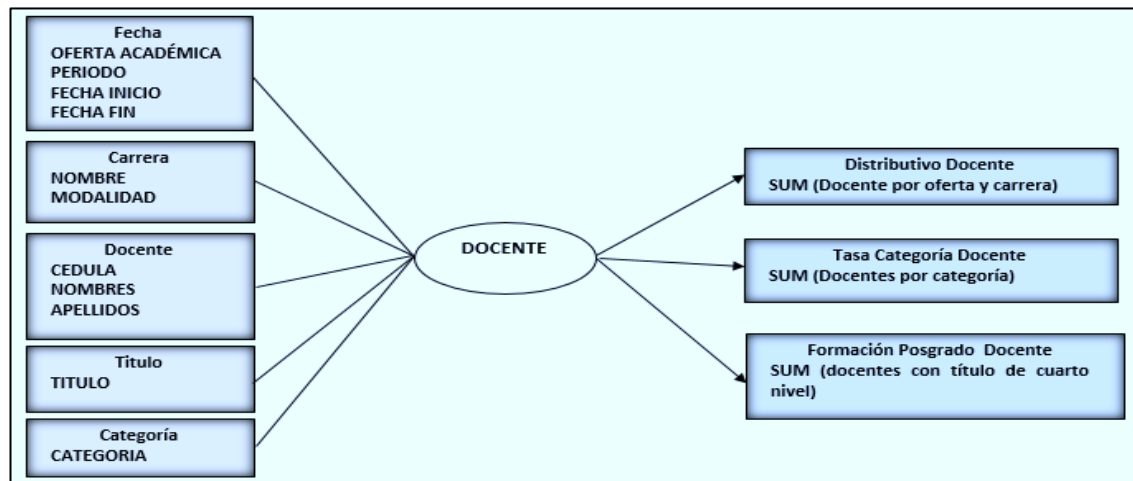


Figura 20 Modelo conceptual ampliado DM_DOCENTES_AEIRNNR.

3.1.3. Etapa 3: Modelo lógico de los Data Marts.

El diseño de los Data Marts para el AEIRNNR, se lo realizó enfocándose en las dimensiones e indicadores definidos anteriormente, se utilizó las técnicas de modelamiento dimensional propuestas por Ralph Kimball, poniendo en práctica el concepto de claves subrogadas (SK, Subrogate Keys) para cargar cada una de las dimensiones. Las claves subrogadas son números enteros generados al aplicar la técnica SCD tipo 2, éstas claves nos ayudarán a mejorar el rendimiento en cada una de las juntas entre dimensiones y tablas de hechos, razón por la cual se añadió a cada dimensión los campos (VALIDO_DESDE, VALIDO_HASTA, VERSION), que se utilizarán al momento de aplicar la técnica SCD tipo 2, la misma que nos permitirá llevar un historial de cambios que puedan darse en las dimensiones de cada Data Mart.

En el Data Mart perteneciente a docentes, se aplicó el concepto de tabla de hechos sin hechos, lo que implica que la relación de la dimensión docente que define la llave de la tabla de hechos, por si sola se refiere a la ocurrencia de un evento (indicadores docentes). Además, se utilizó un esquema de tablas de hechos tipo estrella (Star Scheme), por las razones expuestas a continuación:

- Permite optimizar los tiempos de respuesta ante las consultas de los usuarios.
- Es soportada por la plataforma Pentaho, por lo que los metadatos son fáciles de documentar y mantener.
- Diseño fácilmente modificable.
- Simplifica el análisis.

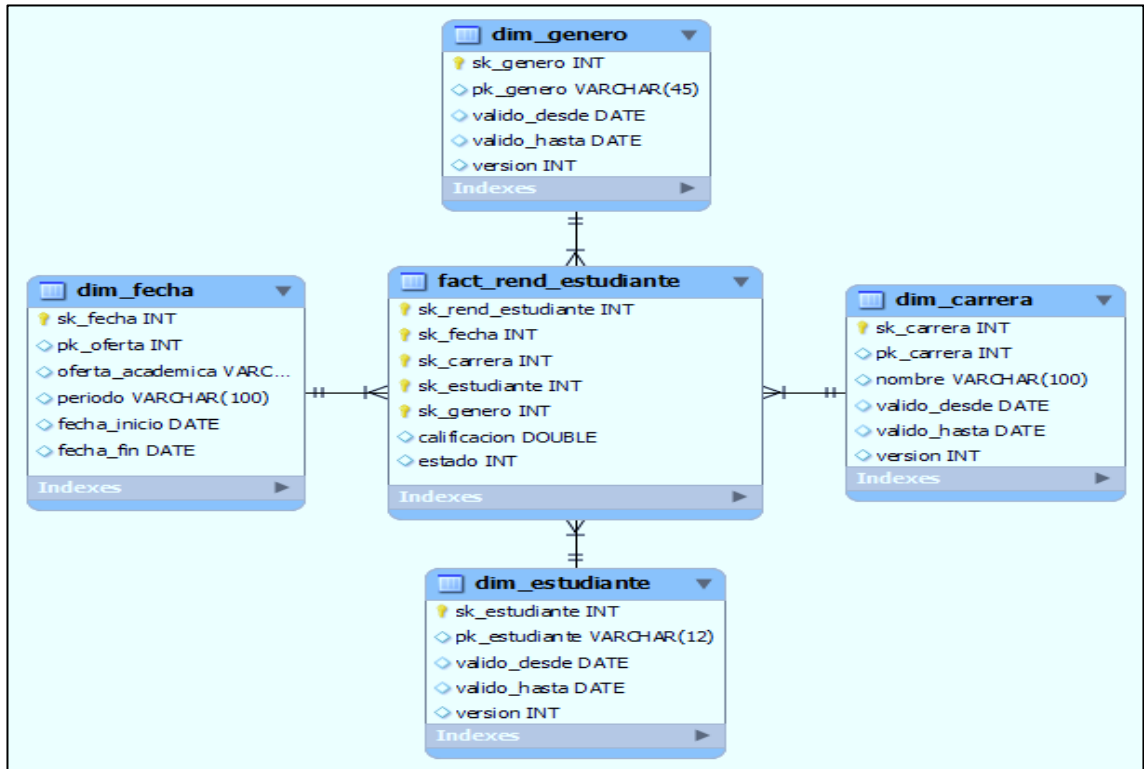


Figura 21 Modelo de datos dimensional DM_ESTUDIANTES_AEIRNNR.

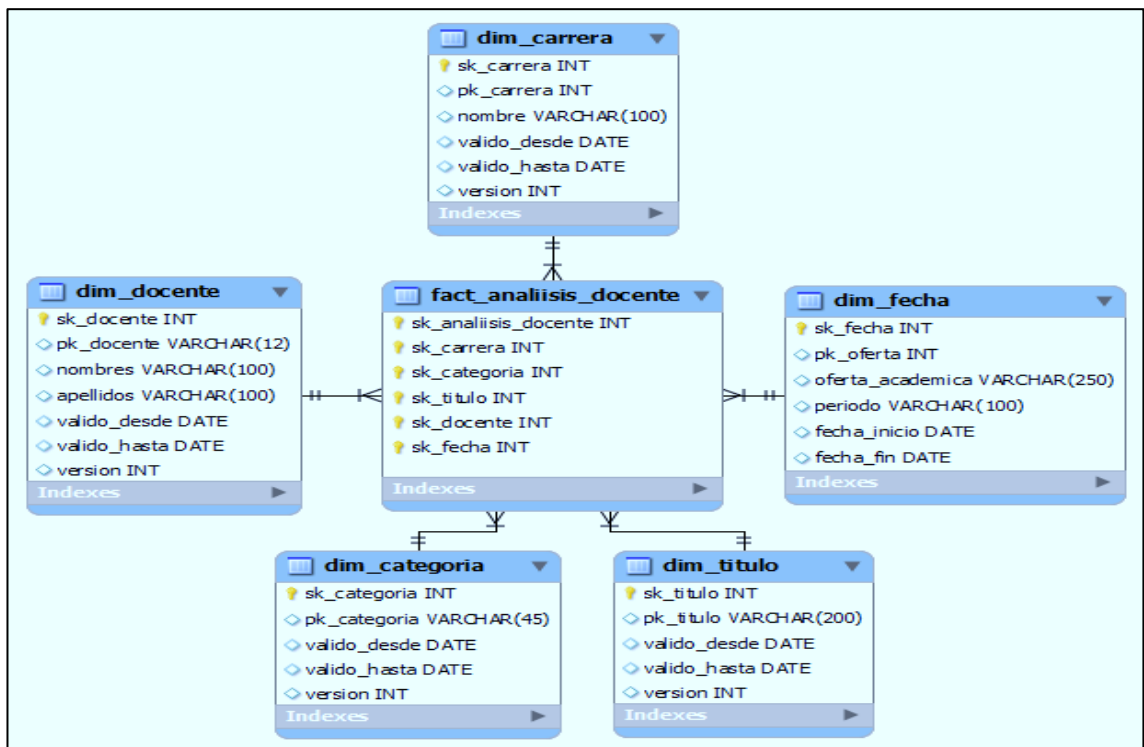


Figura 22 Modelo de datos dimensional DM_DOCENTES_AEIRNNR.

3.1.3.1. Tablas de hechos y tablas de dimensiones.

Cada esquema en estrella definido anteriormente, está conformado por una tabla de hechos denominada FACT_ANALÍISIS_DOCENTE y FACT_REND_ESTUDIANTE respectivamente, todas ellas con sus distintas dimensiones, las mismas que serán explicadas de manera general mediante un diccionario de datos (Véase Anexo 3).

3.1.4. Etapa 4: Integración de datos (Procesos ETL).

3.1.4.1. Arquitectura de flujo de datos.

Para los procesos de integración de datos, se utilizó una arquitectura de flujo de datos simple, dado que se dispone de un único OLTP y principalmente porque se adapta a la arquitectura de Ralph Kimball definida anteriormente, adicional a esto, nos permite aplicar técnicas de calidad de datos y almacenar los metadatos de los procesos ETL para su posterior reutilización, todo este proceso se llevó a cabo mediante la herramienta Pentaho Data Integration.

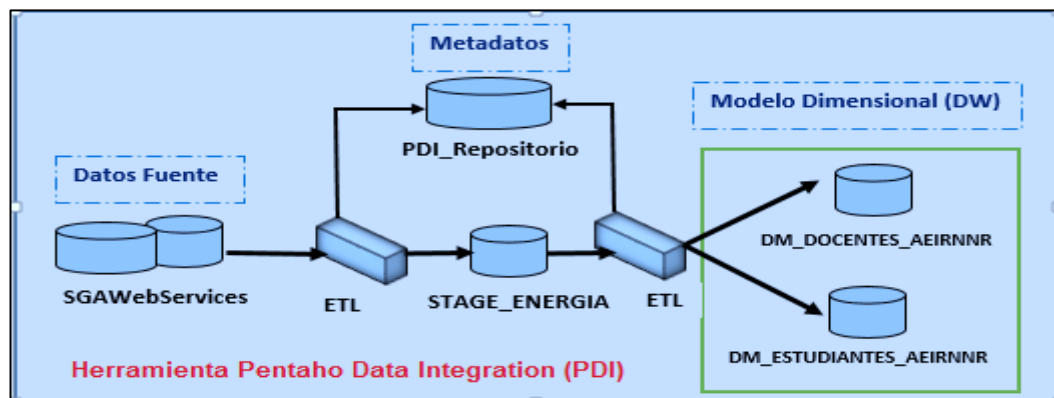


Figura 23 Arquitectura de flujo de datos con PDI.

El proceso consiste en extraer los datos del SGASWebServices, utilizando los métodos del mismo, para luego ser almacenados en la base de datos STAGE_ENERGIA, siendo un proceso bastante complejo porque los datos no residen en un SGBD relacional, además en la salida de cada método utilizado se tendrán que realizar cambios de formato y/o volcado de información, a partir de alguna herramienta específica, en este caso Pentaho Data Integration dispone del componente JSON Input, el mismo que nos sirvió de apoyo en este proceso, ya que la salida de cada método es devuelta en formato JSON (JavaScript Object Notation). Para tener una visión más clara de la información devuelta por cada método del SGASWebServices, a continuación se muestra la salida que nos devuelve el método sgaws_modulos_carrera.



```
<result>
["Pregrado Septiembre 2014 - Febrero 2015", "Ingenier\u00eda en Sistemas", "presencial",
["id_paralelo,seccion,numero_modulo,nombre_paralelo,id_modulo"],
[[18826, "matutino", "9", "B", 61], [18825, "matutino", "9", "A", 61],
[18824, "matutino", "7", "B", 59], [18823, "matutino", "7", "A", 59],
[18822, "matutino", "4", "A", 56], [18821, "matutino", "2", "D", 54],
[18820, "matutino", "2", "C", 54], [18819, "matutino", "2", "B", 54],
[18818, "matutino", "2", "A", 54]]]
</result>
```

Figura 24 Salida de información SGAWebServices.

Como se puede apreciar en la Fig. 24, la salida de información que nos devuelve cada método contiene los siguientes elementos:

- **Objeto:** conjunto desordenado de pares **nombre/valor**. Un objeto comienza con llave de apertura (**{**) y termina con llave de cierre (**}**). Cada nombre es seguido por dos puntos (**:**) y los pares **nombre/valor** están separados por coma (**,**) [28].
- **Arreglo:** Un arreglo es una colección de valores, comienza con un corchete izquierdo (**[**) y termina con corchete derecho (**]**). Los valores se separan por coma (**,**). Un valor puede ser una cadena de caracteres con comillas dobles, un número, un valor booleano (true o false), un valor nulo (null), un objeto o un arreglo [28].

La salida de cada método del SGAWebServices, sigue la misma estructura mostrada en la Fig. 24, aunque varían unos de otros pero no en gran medida, de allí la necesidad de realizar cambios de formato y/o volcado de información, a partir de la configuración del salto Web Services Lookup y JSON Input, que son la base para extraer los datos del SGAWebServices.

Una vez cargada la base de datos STAGE_ENERGIA, que es la que actuará como base de datos temporal, se podrá realizar las operaciones de transformación de datos, limpieza y carga a cada uno de los Data Marts. El modelo relacional desarrollado para el STAGE_ENERGIA se muestra a continuación en la Fig. 25.

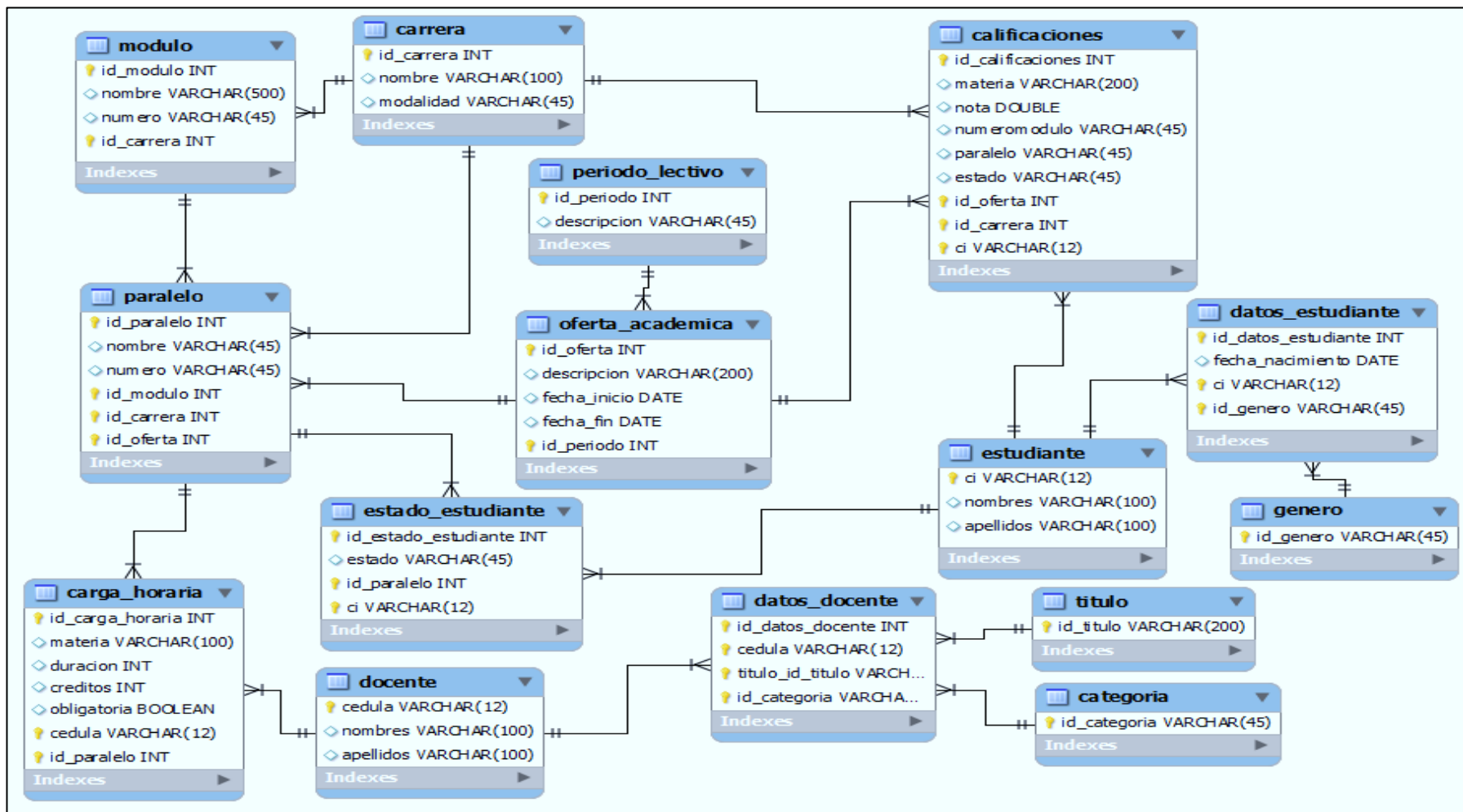


Figura 25 Modelo Relacional STAGE_ENERGIA.

En Pentaho Data Integration, los procesos ETL son gestionados por transformaciones, saltos (steps), pasos (hops), trabajos (jobs) y otras estructuras.

Una transformación, es una herramienta que permite realizar una variedad de tareas con data moviéndolos de un lugar a otro, su función principal es extraer data de fuentes de datos, transformar la data de una representación a otra y cargar data en fuentes de salida (output sources), es decir realiza los pasos de extracción (E), transformación (T) y carga (L).

Una transformación consiste de un número separado de acciones llamados steps (pasos), cada uno diseñado para una función específica. La representación gráfica del flujo de datos entre los steps son llamados hops (saltos). Los hops son usados para transportar data de un step a otro.

Un job, es un componente que crea una secuencia de actividades y brinda un orden de ejecución, por lo tanto los jobs son usados para un control de flujo y por lo general consiste de una serie de transformaciones.

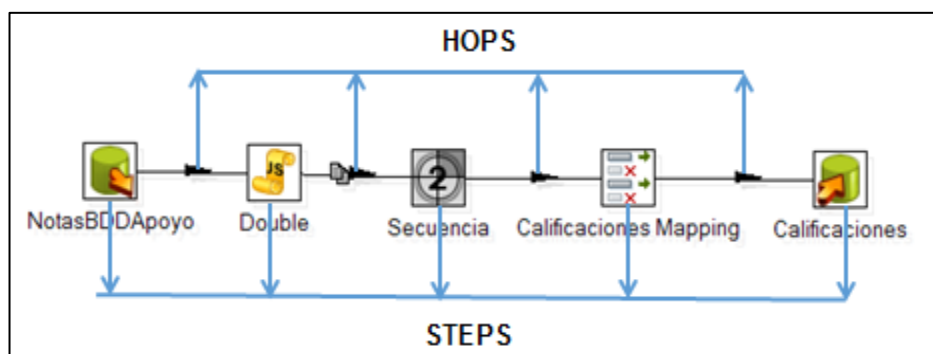


Figura 26 Ejemplo de Transformación en PDI.

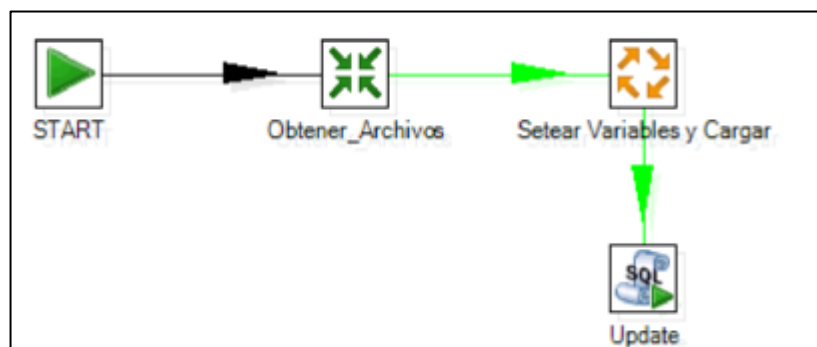


Figura 27 Ejemplo de trabajo (Job) con PDI.



Continuando con la arquitectura de flujo de datos definida anteriormente, para el desarrollo del proyecto se instaló el SGBD (Sistema de Gestión de Bases de Datos) PostgreSQL y la herramienta de gestión pgAdmin III, las mismas que sirvieron de apoyo en la creación de los siguientes esquemas:

- **PDI_REPOSITORIO:** permitió almacenar los procesos de carga y los metadatos manejados por la herramienta Pentaho Data Integration, es decir es un conjunto de tablas que contiene los objetos creados por el diseñador y los objetos predefinidos del sistema, los metadatos de las fuentes y destinos, y las reglas de transformación.
- **STAGE_ENERGIA:** base de datos de almacenamiento temporal, que nos permitió realizar operaciones de transformación y limpieza de datos, previos al ingreso de información a cada uno de los Data Marts.
- **DM_DOCENTES_AEIRNNR:** bodega de datos en donde se almacenó la información sumariada relacionada con la parte docente del AEIRNNR.
- **DM_ESTUDIANTES_AEIRNNR:** bodega de datos en donde se almacenó la información sumariada relacionada con la parte de estudiantes del AEIRNNR, la información de ambos Data Marts debe estar depurada, limpia y lista para poder ser utilizada por los otros módulos de Pentaho (Report Designer, CDE, etc.).

Además, en Pentaho Data Integration es necesario crear las conexiones a los almacenes de datos que actuarán como fuente y destino. En el *Anexo 4*, se muestra el proceso que se siguió para crear la conexión al STAGE_ENERGIA, que es la base de datos que actuará como almacenamiento temporal, este procedimiento es el mismo que se sigue para realizar la conexión a los diferentes Schemas, variando solo los nombres de cada base de datos.

3.1.4.2. Implementación de flujos de carga.

La implementación de flujos de carga se la realizó mediante cuatro flujos de trabajo (Jobs), el primero que nos ayudó a poblar el Stage o almacenamiento temporal (STAGE_ENERGIA), el segundo y tercero que nos ayudaron a poblar los Data Marts (DM_DOCENTES_AEIRNNR, DM_ESTUDIANTES_AEIRNNR) a partir de datos contenidos en el STAGE_ENERGIA, y finalmente se unió cada uno de estos flujos de trabajo mediante una secuencia de ejecución de los mismos (Job).



3.1.4.2.1. Flujo de trabajo STAGE_ENERGIA.

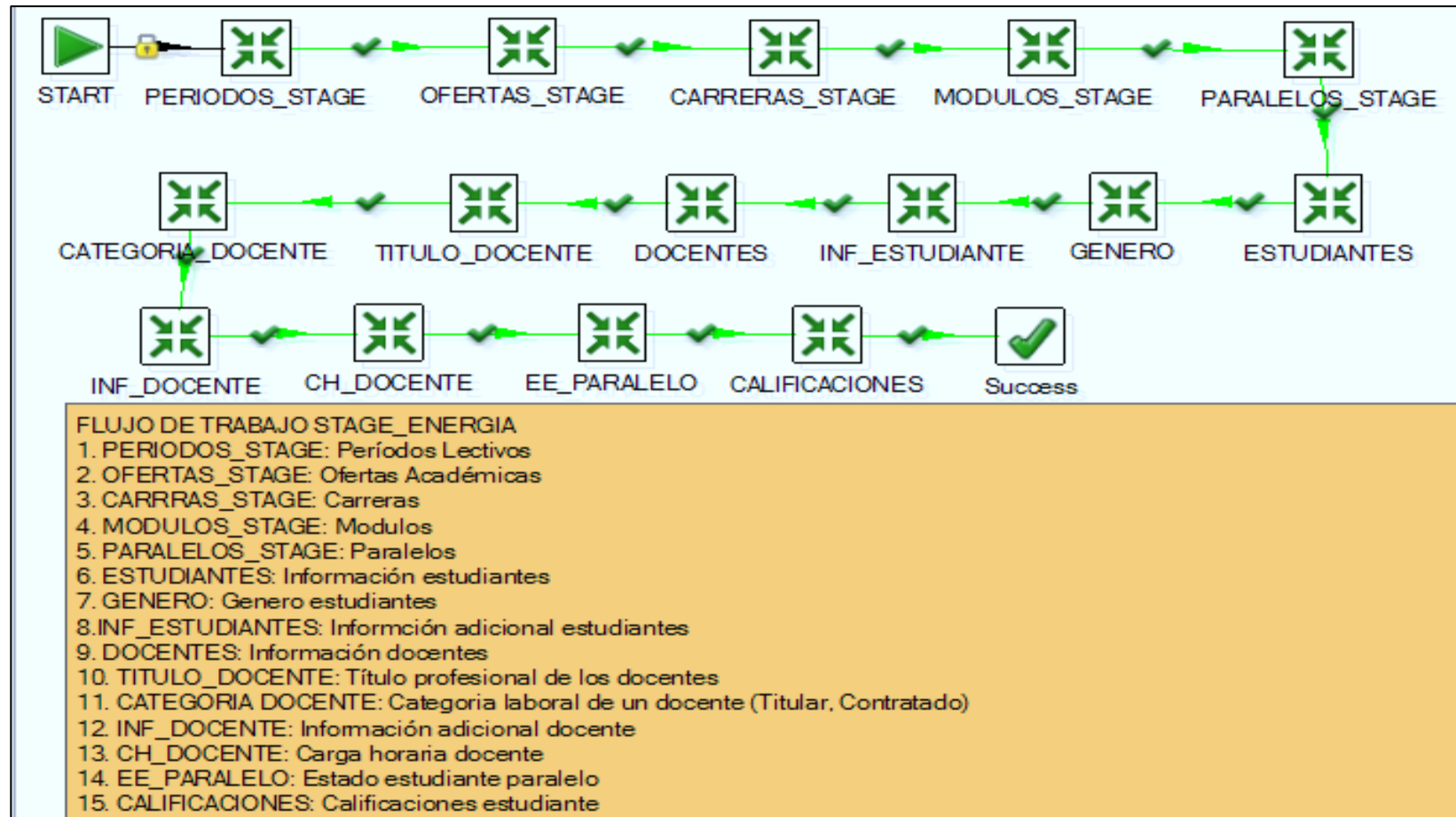


Figura 28 Job STAGE_ENERGIA.



El flujo de trabajo STAGE_ENERGIA, contiene la secuencia de transformaciones que permiten extraer la información del SGASWebServices, transformarla y cargarla en el almacén de datos temporal, la información devuelta por el SGASWebServices, como se mencionó con anterioridad sigue la estructura mostrada en la Fig. 24, por tal motivo se puede deducir que los pasos que intervienen en las diferentes transformaciones construidas para poblar el almacén de datos temporal, varían muy poco en su configuración. En el Anexo 5, se detalla de manera general las configuraciones realizadas a los pasos que intervienen en la transformación CARGA_OFERTAS_STAGE, contenida dentro del trabajo STAGE_ENERGIA.

Se debe mencionar que la ejecución de cada una de las transformaciones creadas para poblar el almacén de datos temporal, se las realizó de manera secuencial, ya que unos flujos dependen de otros, por ejemplo, para ejecutar la transformación CARGA_OFERTAS_STAGE, dentro del cual se trabaja con el método sgaws_ofertas_academicas, que recibe como parámetro la clave primaria de cada periodo lectivo, necesitamos que el flujo CARGA_PERIODOS_STAGE ya este ejecutado, porque este alimenta de datos la tabla periodos del STAGE_ENERGIA, que es de donde se extraerá la clave primaria de cada periodo para poder ejecutar dicha transformación.

3.1.4.2.2. Caracterización de datos.

Después de alimentar de datos el STAGE_ENERGIA, se realizó el proceso de caracterización de datos, mediante el plugging de Pentaho Data Integration denominado DataCleaner Profiling.

La caracterización de datos, es el proceso de recopilación de información sobre los datos existentes en nuestros orígenes de información. Esta información, es de gran utilidad en el diseño de los procesos ETL para cargar los Data Marts, también es parte importante de cualquier iniciativa de calidad de datos, ya que antes de que la calidad de estos se pueda mejorar, habrá que establecer cuál es el estado actual de los datos, y para ello nos sirvió esta técnica. Este proceso se lo realizó al nivel de **Column Profile** sobre el STAGE_ENERGIA, es decir se analizó cada una de las columnas de las tablas del mismo, permitiéndonos encontrar errores como:

- Errores de integridad entre valores de diferentes tablas (ejemplo: nombres de docentes con espacios en blanco al inicio o al final del mismo).



- Valores fuera de rangos posibles (cadenas ingresadas como notas de una materia).
- Valores de columnas con errores tipográficos (ejemplo: se encontró nombres de materias mal ingresadas).
- Datos mal ingresados (ejemplo: nombres de estudiantes, docentes en minúsculas).
- Falta de integridad referencial.

El error más significativo al momento de realizar la caracterización de datos sobre el STAGE_ENERGIA, se lo encontró dentro de la tabla calificaciones, exclusivamente dentro de la columna nombre (nombre de la materia) y nota (nota perteneciente a la materia), se detectó que el SGASWebServices nos devuelve información errónea del tipo: **nombre:** U4, **nota:** ESTÁTICA, perteneciendo la mayoría de estos registros a la carrera de Ingeniería en Electrónica y Electromecánica, por lo que se desechó este tipo de información, debido a que el nombre de las materias devueltas no está completo, pero sobre todo porque el valor de la nota no es un valor numérico, por tal motivo se los ha considerado no aptos para el análisis. A continuación en la Fig. 29, se muestra el error detectado mediante la caracterización de datos:

idnotas	nombre	nota
500	U4	ESTATICA
512	U4	DINAMICA
522	U4	Termodinámica
1094	U4	Practica de circuitos electronicos analogicos
1321	U4	Practica de circuitos electronicos analogicos
1507	U4	Practica de circuitos electronicos analogicos
3755	U4	HERRAMIENTA CAD I
4080	U4	ESTATICA
4511	U4	Practica de circuitos electronicos analogicos
5944	U4	Practica de circuitos electronicos analogicos
6108	U4	HERRAMIENTA CAD I
6144	U4	TERMODINAMICA
6152	U4	Tecnología de los Metales
6162	U4	Transferencia de Calor
13913	U4	Practica de circuitos electronicos analogicos
27632	U4	TECNOLOGIA DE LOS METALES
27728	U4	Practica de circuitos electronicos analogicos
31250	U4	ESTATICA
36568	U4	ESTATICA

Figura 29 Error detectado mediante Caracterización de datos.



3.1.4.2.3. Flujo de trabajo UNIR_TRANSF_ESTUDIANTE.

El flujo de trabajo UNIR_TRANSF_ESTUDIANTE, contiene los procesos ETL que permiten extraer los datos del STAGE_ENERGIA, transformarlos y depositarlos limpios en las tablas del esquema DM_ESTUDIANTES_AEIRNNR, después de realizada la caracterización de datos. Los flujos o transformaciones construidos para poblar las dimensiones y los hechos del DM_ESTUDIANTES_AEIRNNR (DIM_CARRERA_ESTUDIANTE, DIM_GENERO_ESTUDIANTE, DIM_DM_ESTUDIANTE, DIM_FECHA_ESTUDIANTE, HECHOS_ESTUDIANTE), siguen la misma estructura, excepto el flujo HECHOS_ESTUDIANTE y DIM_FECHA_ESTUDIANTE, que siguen una estructura diferente. En el Anexo 6, se detalla de manera general las transformaciones HECHOS_ESTUDIANTE, DIM_FECHA_ESTUDIANTE y DIM_CARRERA_ESTUDIANTE, que están contenidas dentro de este flujo de trabajo.

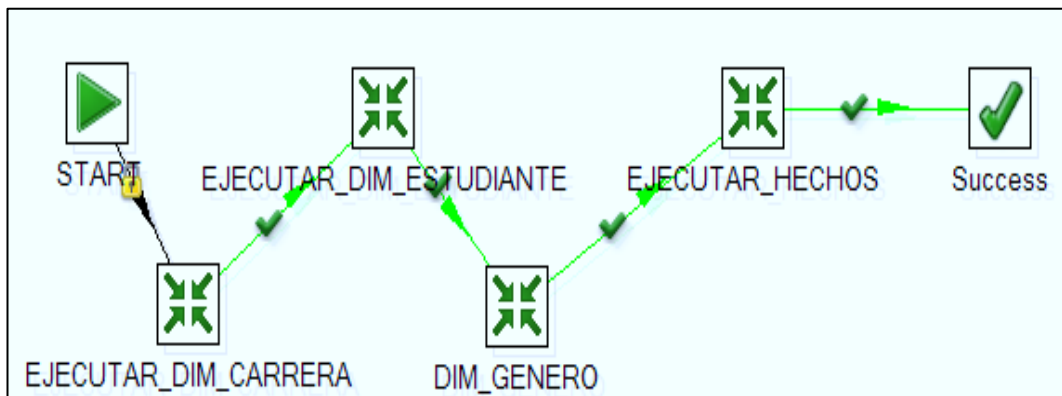


Figura 30 Job UNIR_TRANSF_ESTUDIANTE.

3.1.4.2.4. Flujo de trabajo UNIR_TRANSF_DOCENTE.

El flujo de trabajo UNIR_TRANSF_DOCENTE, contiene los procesos ETL que permiten extraer los datos del STAGE_ENERGIA, transformarlos y depositarlos limpios en las tablas del esquema DM_DOCENTES_AEIRNNR.

Los flujos de carga o transformaciones construidos para poblar las dimensiones y los hechos del DM_ESTUDIANTES_AEIRNNR (DIM_CARRERA_DOCENTE, DIM_DM_DOCENTE, CARGAR_DIM_CATEGORIA, CARGAR_DIM_TITULO, HECHOS_DOCENTE, DIM_FECHA_DOCENTE) siguen la misma estructura del flujo de trabajo anterior, por lo que no se entrara en detalle.

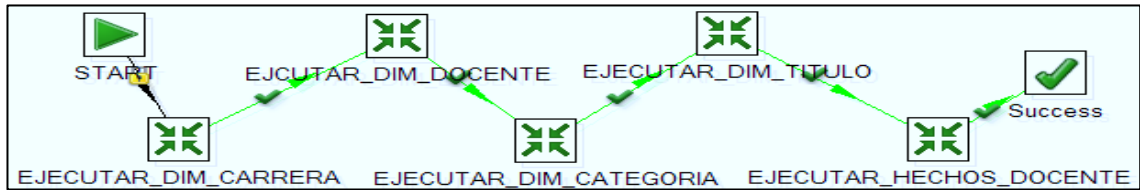


Figura 31 Job UNIR_TRANSF_DOCENTE.

3.1.4.2.5. Flujo de trabajo UNIR_DM_STAGE.

Los flujos de trabajo anteriores, contienen cada una de las transformaciones para poblar el STAGE_ENERGIA y cada uno de los Data Marts, una vez terminadas las configuraciones de estos flujos, se estableció un orden de ejecución de los mismos mediante el flujo de trabajo UNIR_DM_STAGE, que es el que contiene la secuencia de ejecución de dichos flujos. El flujo UNIR_DM_STAGE, tiene la particularidad de ejecutar una serie de trabajos, y estos trabajos a su vez contienen una serie de transformaciones. A continuación en la Fig. 32, se muestra el orden de ejecución del mismo.

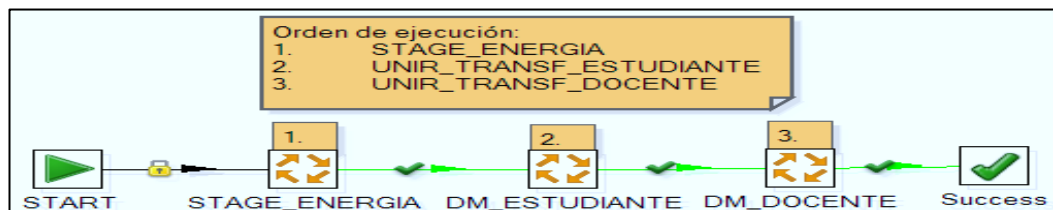


Figura 32 Job UNIR_DM_STAGE.

3.1.4.2.6. Actualización del Data Warehouse.

El problema de la frecuencia de carga de los Data Marts que conforman el Data Warehouse del área, se lo ha definido de manera mensual, dado que las transacciones del SGAWebServices (carga de notas, estudiantes, paralelos, etc.), se llevan a cabo al inicio y al finalizar cada oferta académica (semestre).

Para programar el proceso de actualización, se utilizó la herramienta Kitchen de Pentaho Data Integration, el mismo que nos permite ejecutar trabajos diseñados con Spoon (diseñador gráfico de PDI), siempre y cuando estén almacenados como archivos XML o en el repositorio de base de datos. Generalmente los trabajos se programan en modo por lotes (archivos .bat), para ser ejecutados en intervalos de tiempo regulares, mediante el comando AT de Windows o el programador de tareas, los mismos que permiten programar la ejecución automática de aplicaciones en un computador a una hora y fecha específicas, en nuestro caso particular se ha utilizado



el programador de tareas de Windows. Los archivos .bat creados para automatizar la ejecución del trabajo UNIR_DM_STAGE, se muestran a continuación:

- **ejecutor_dm_docente.bat**
cd C:\pentaho\data-integration
- **dm_docente.bat**
kitchen.bat /rep: REPOSITORIO /user: admin /pass: admin /dir:/
/job:UNIR_DM_STAGE /level:Basic
- **actualizador_dm.bat**
Call C:\pentaho\data-integration\ejecutor_dm_docente.bat
Call C:\pentaho\data-integration\dm_docente.bat

Dentro del archivo **ejecutor_dm_docente.bat**, se accede a la ruta de instalación de Pentaho Data Integration mediante el comando cd, y a través del archivo **dm_docente.bat**, se ejecuta el archivo kitchen.bat, contenido dentro de la carpeta de instalación de PDI, el mismo que recibe como parámetros el nombre del repositorio, el usuario, contraseña, dirección y el nombre del trabajo a ejecutar.

Una vez establecidos los archivos que nos permiten automatizar la ejecución del flujo de trabajo UNIR_DM_STAGE, se realizó una secuencia de ejecución de los mismos mediante el archivo **actualizador_dm.bat**, finalmente se creó una tarea programada mediante el programador de tareas de Windows, en donde se especificó que el archivo **actualizador_dm.bat**, se deberá ejecutar el día 25 de cada mes.

3.2. Aplicación del DW para el análisis de la información académica del AEIRNNR (Extracción de la información).

Una vez construido el DW para el análisis de la información académica del AEIRNNR, es necesario mostrar la información generada por el mismo, para que esta sea analizada por el usuario final a través de cubos OLAP, reportes, etc.

3.2.1. Creación de cubos multidimensionales (Reportes Dinámicos).

Para poder realizar análisis OLAP, se utilizó el módulo Pentaho Schema Workbench, los cubos que utiliza este servidor son llamados esquemas y son archivos XML que definen la estructura de los mismos (dimensiones, jerarquías, niveles, hechos, medidas), además contienen información de cómo se realizan los mapeos entre estas estructuras y las tablas de cada Data Mart, la principal ventaja del por qué utilizar análisis OLAP, radica en que las consultas son respondidas con gran performance, minimizando al máximo el tiempo que se hubiese incurrido en realizar dicha consulta,



sobre una base de datos transaccional. A continuación, se presenta la estructura general de los cubos utilizados dentro del presente proyecto, la estructura de cada cubo se detalla como un árbol en la cual a partir del nodo padre Cubo, se desprenden cada uno de los componentes que lo definen, en el Anexo 8, se explica de manera general el proceso para crear cubos OLAP con Pentaho Schema Workbench.

3.2.1.1. Cubo Docentes.

El cubo para el análisis de la información de docentes, se lo ha denominado CUBO_PROFESORES, y este nos permitió visualizar el total de docentes que existen en una determinada oferta académica, así como la distribución de los mismos dentro de cada carrera y clasificarlos por categoría laboral. El CUBO_PROFESORES, contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último las medidas calculadas necesarias para la obtención de los indicadores antes expuestos.

Desplegando la estructura correspondiente al diseño de la dimensión DIM_FECHA, se puede observar la representación de una jerarquía interna de dos niveles. Estos son PERIODOS, OFERTAS, el resto de dimensiones solamente contienen un nivel.

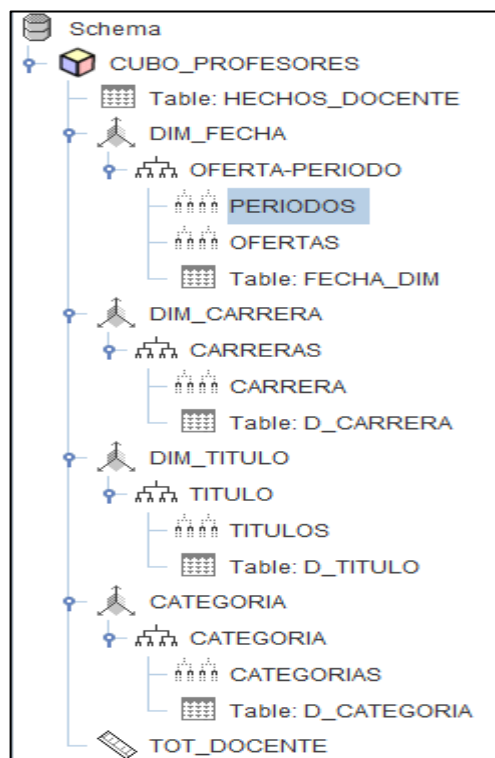


Figura 33 Representación CUBO_PROFESORES.



3.2.1.2. Cubo Estudiantes.

El cubo para el análisis de la información de estudiantes, se lo ha denominado CUBO_RENDIMIENTO, y este nos permitió visualizar el total de estudiantes y el promedio de notas que existen en una determinada oferta académica y carrera. Este cubo contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último las medidas calculadas, necesarias para la obtención de los indicadores académicos. El CUBO_RENDIMIENTO, al igual que el anterior contiene una jerarquía interna de dos niveles dentro de la dimensión DIM_FECHA (PERIODOS, OFERTAS), mientras que el resto de dimensiones contienen únicamente un nivel.

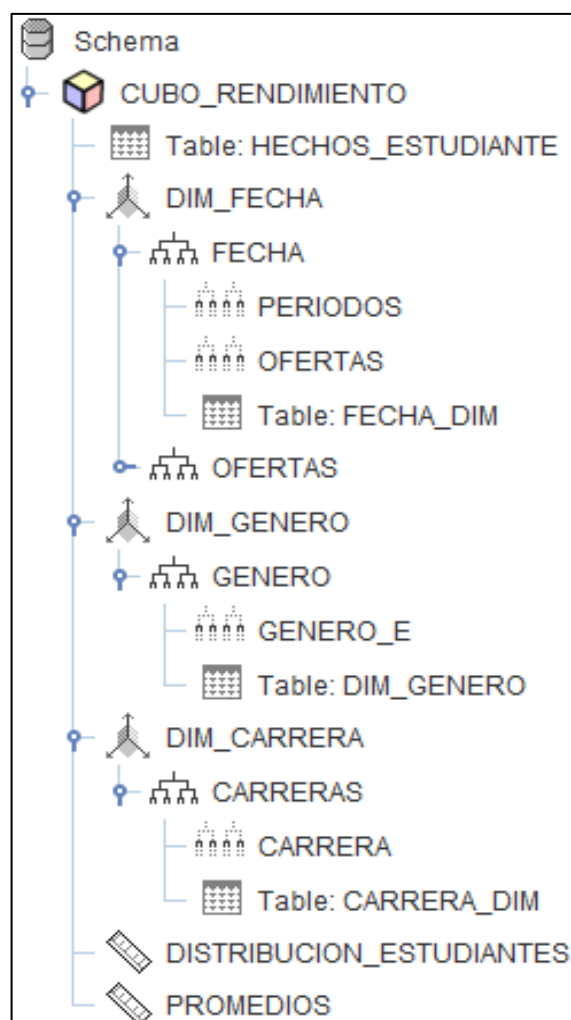


Figura 34 Representación CUBO_RENDIMIENTO.



3.2.2. Generación de reportes personalizados.

El componente de Pentaho utilizado para la creación de reportes con los indicadores académicos del AEIRNNR, es Pentaho Report Designer, el mismo que mediante conexión a los Data Marts creados y consultas SQL, nos permitió dar solución a dichos indicadores. En el *Anexo 7*, se muestra la construcción de reportes con Pentaho Report Designer.

Una definición general de un reporte, consiste en un conjunto de secciones que definen la disposición y contenido de la información dentro de éste. Los reportes generados con Pentaho Report Designer, nos han permitido dar solución a los indicadores: Formación de Posgrado, Tasa de profesores por Categoría, Tasa de aprobados y reprobados, de manera general estos reportes están compuesto por tres partes:

- La cabecera: contiene el nombre de la institución.
- El cuerpo: genera información académica como el total de alumnos matriculados por oferta, total de alumnos aprobados y reprobados, información representada en gráficos estadísticos, etc.
- Filtro: permite visualizar el reporte por oferta académica y en muchos casos por carrera (reportes parametrizados).

3.2.3. Presentación de información al usuario final.

Una vez que se ha diseñado los cubos OLAP (Schema Worbench), y los reportes (Report Designer), Pentaho ofrece un módulo adicional denominado Pentaho BI Server, el mismo que mediante un entorno web, nos permitió mostrar la información al usuario final, a través de la visualización de los cubos y reportes generados con las herramientas antes mencionadas.

Para la visualización de cubos OLAP, Pentaho BI Server posee el visor denominado JPivot, pero la visualización que este genera no es muy amigable al usuario, por lo que se instaló el plugin denominado Saiku Analisis (*Vease Anexo 10*), el mismo que nos proporcionó una visión bastante mejorada de los cubos OLAP. Adicional a esto, se realizó la conexión a los Data Marts creados anteriormente, para posteriormente importar los cubos y reportes creados. A continuación se mostrará la visualización de los cubos OLAP y cada uno de los reportes mediante el módulo Pentaho BI Server.



3.2.3.1. Tasa de Aprobados y Reprobados.

El reporte de estudiantes aprobados y reprobados por oferta académica, permite al usuario seleccionar una oferta académica, y en base a esta despliega resultados como: total de alumnos matriculados, alumnos reprobados, alumnos aprobados, para finalmente mostrar mediante gráficos estadísticos el porcentaje que representan estas medidas en cada carrera.

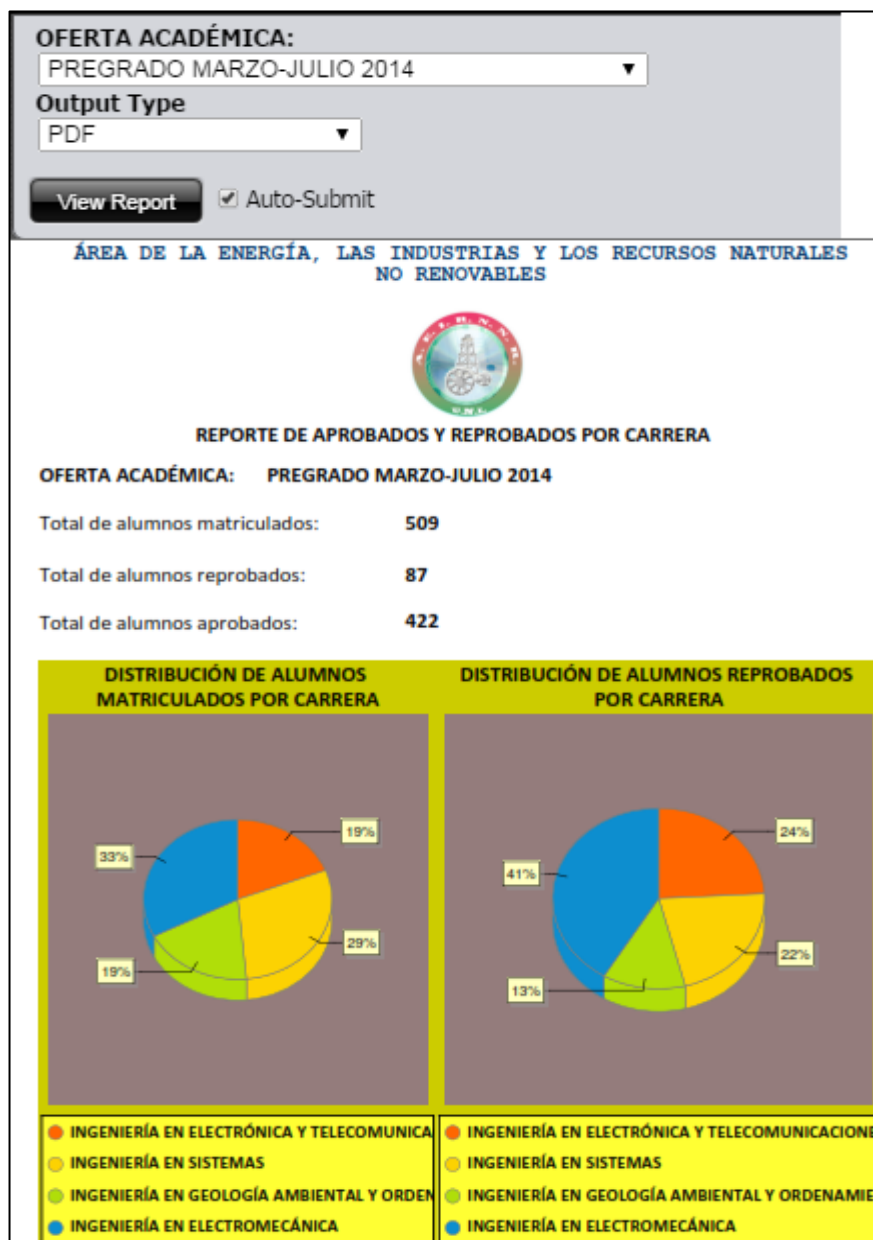


Figura 35 Estudiantes Aprobados y Reprobados por carrera.



3.2.3.2. Tasa de profesores por Categoría.

Para mostrar al usuario la tasa de profesores por categoría en una determinada oferta académica y carrera, se generó un reporte que le permita seleccionar la oferta académica y categoría laboral (titular, contratado) de la cual se desea realizar el análisis, después de seleccionar dichos parámetros se mostrará en pantalla el total de docentes que existen por oferta académica y carrera de la categoría seleccionada, y finalmente esta información se despliega en gráficas estadísticas.

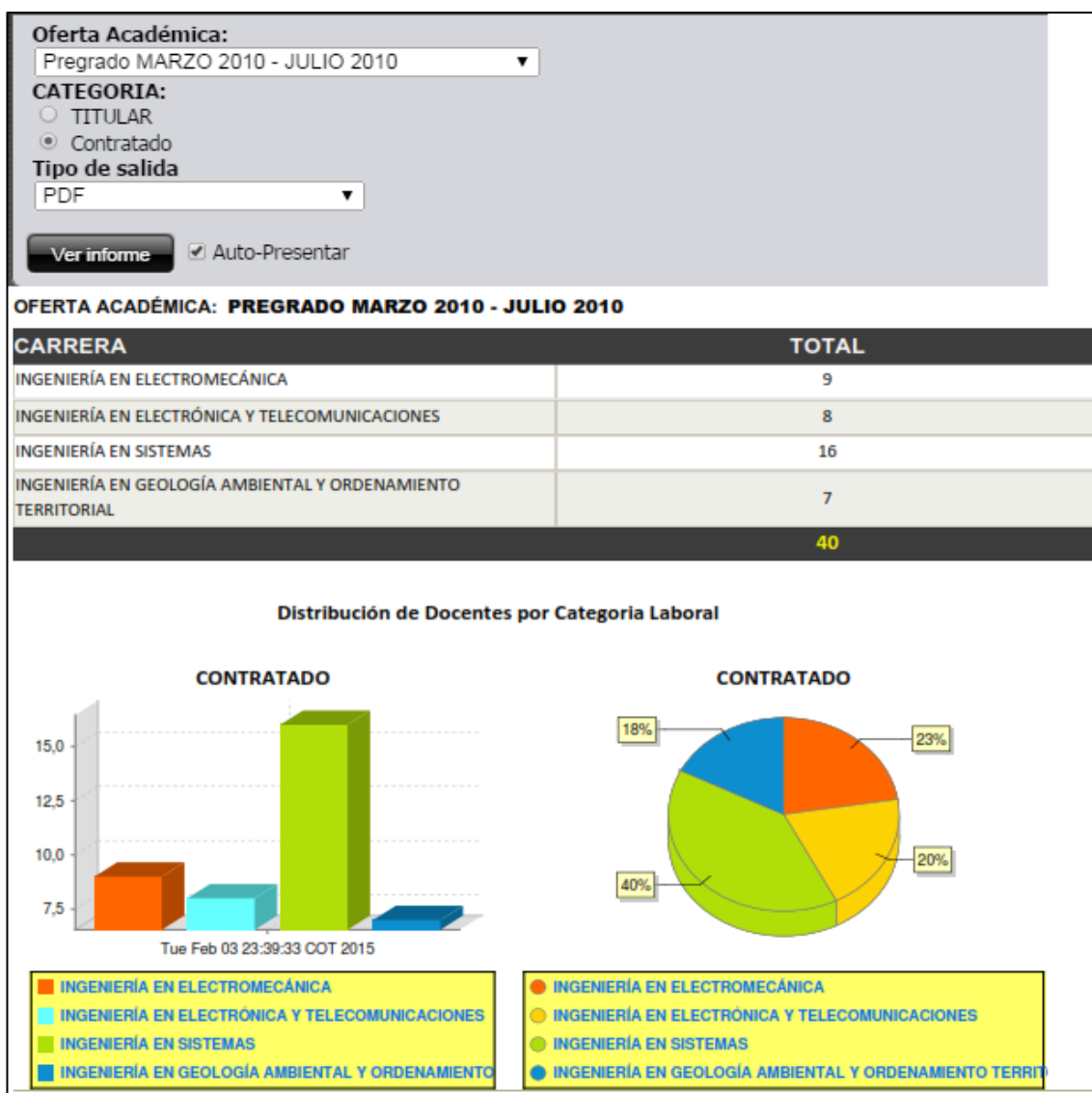


Figura 36 Tasa de profesores por Categoría Laboral.



3.2.3.3. Formación Posgrado Docentes.

El reporte formación posgrado docente, permite al usuario seleccionar una determinada oferta académica y carrera, y en base a estos dos parámetros devuelve los nombres, apellidos y el título profesional de los docentes con formación profesional de cuarto nivel, e información estadística del título profesional con mayor número de docentes.

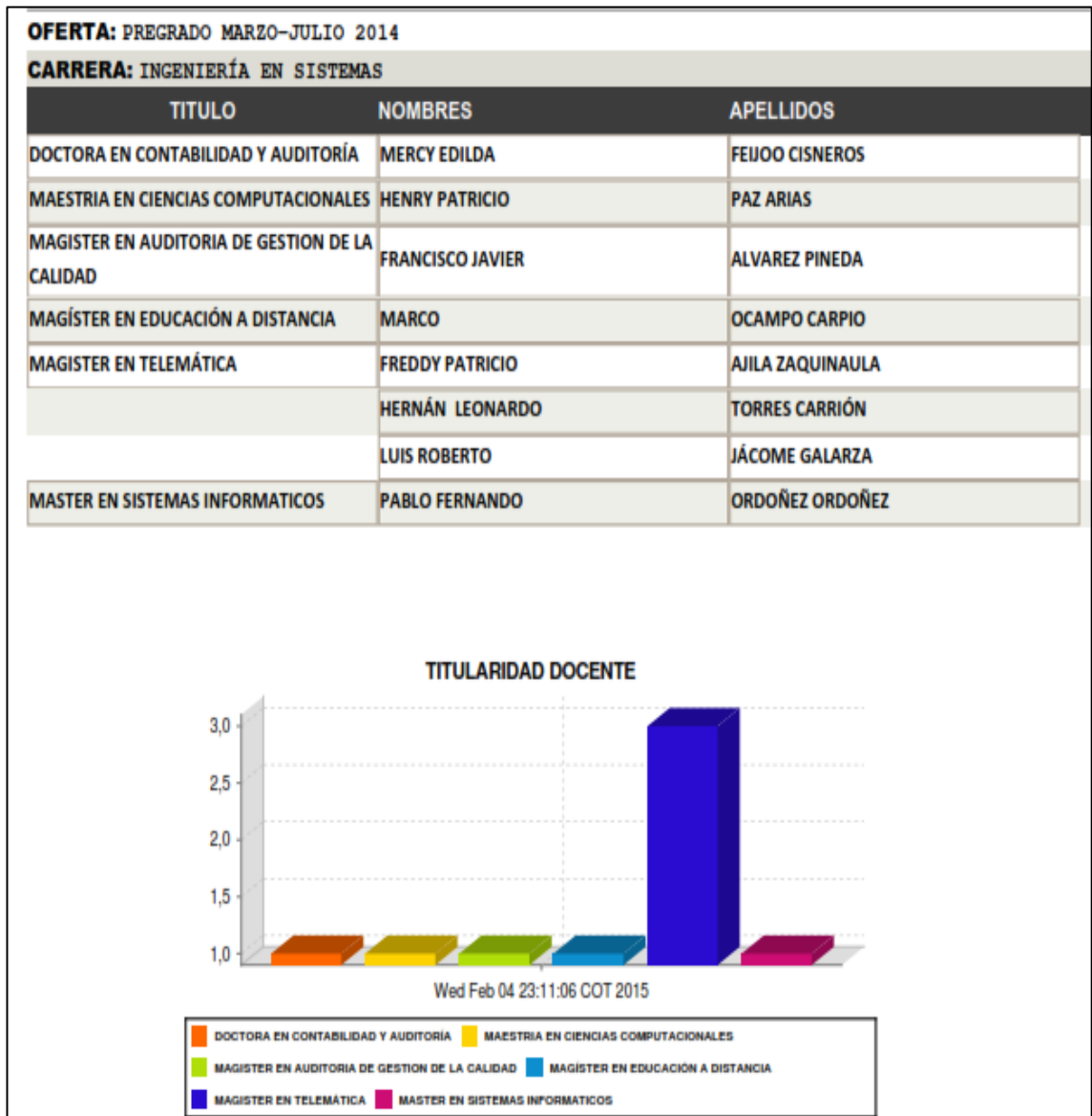


Figura 37 Formación posgrado docente.

3.2.3.4. Total de docentes por Oferta Académica y Carrera.

Para mostrar al usuario el total de docentes por oferta académica, se lo realizó mediante el cubo OLAP de docentes CUBO_PROFESORES, en donde el usuario elige el indicador TOT_DOCENTE, posteriormente selecciona la dimensión DIM_FECHA, y finalmente se genera la información de cada una de las ofertas académicas, pudiendo además desplegar este informe por carrera.



Figura 38 Docentes por Oferta Académica.

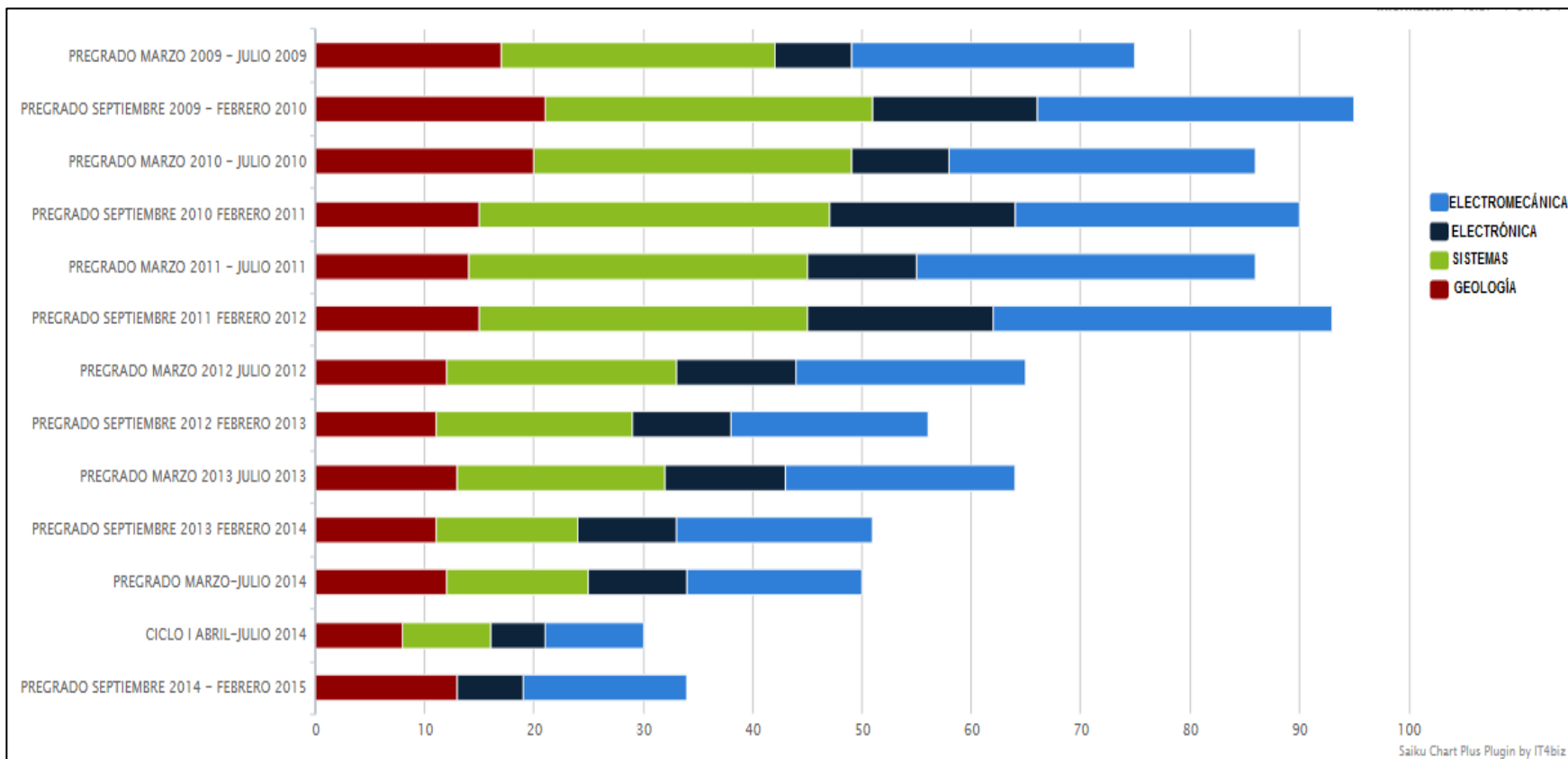


Figura 39 Distribución docente por Oferta Académica y Carrera.

3.2.3.5. Promedio de notas por Oferta Académica y Carrera.

El reporte Promedio de notas por Oferta Académica y Carrera, se lo realizó mediante el cubo OLAP CUBO_RENDIMIENTO, el usuario puede desplegar esta información mediante el indicador PROMEDIOS contenido dentro del mismo cubo, el análisis de esta información se lo puede realizar a través de la dimensión fecha y carrera.

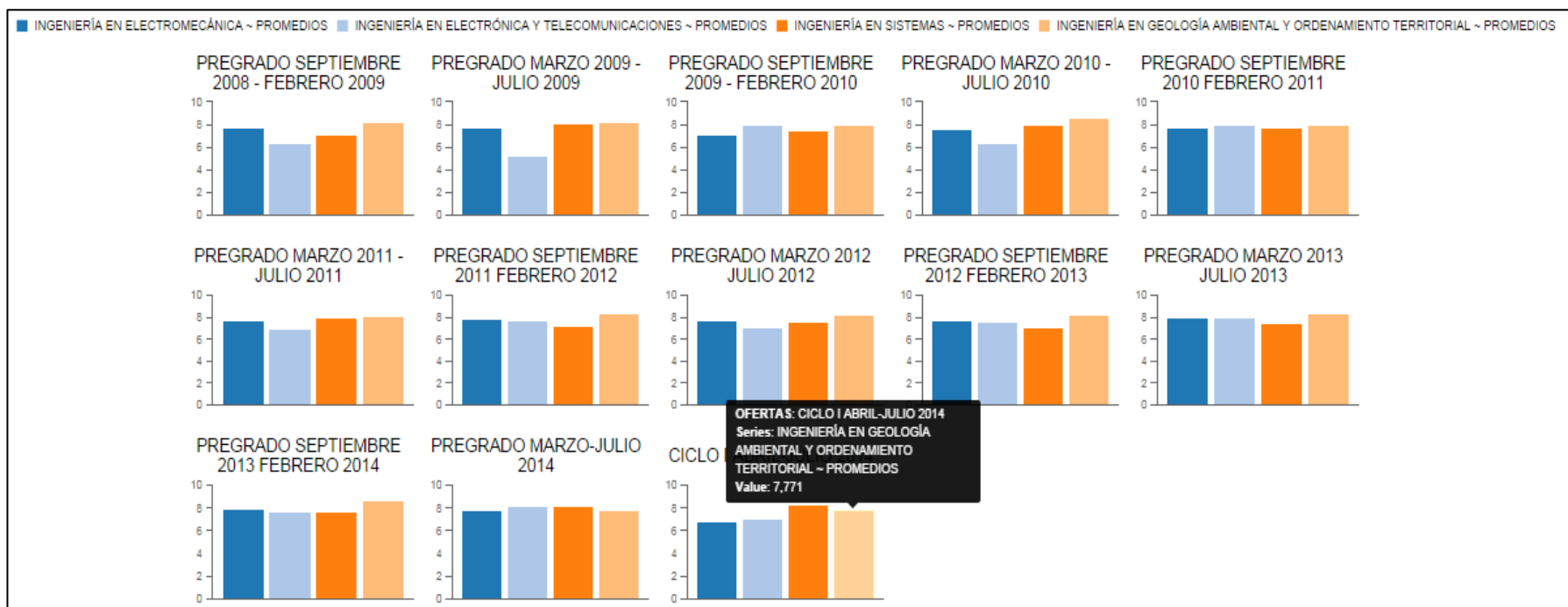


Figura 40 Promedio de notas por Oferta Académica y Carrera.

3.2.3.6. Tasa de estudiantes por Oferta Académica – Distribución estudiantes por Carreras.

El reporte Tasa de estudiantes por Oferta Académica – Distribución estudiantes por Carreras, permite al usuario visualizar en una sola pantalla, el total de estudiantes que existen en una determinada oferta académica y como estos se encuentran distribuidos en las diferentes carreras, toda esta información la presenta en gráficos estadísticos.

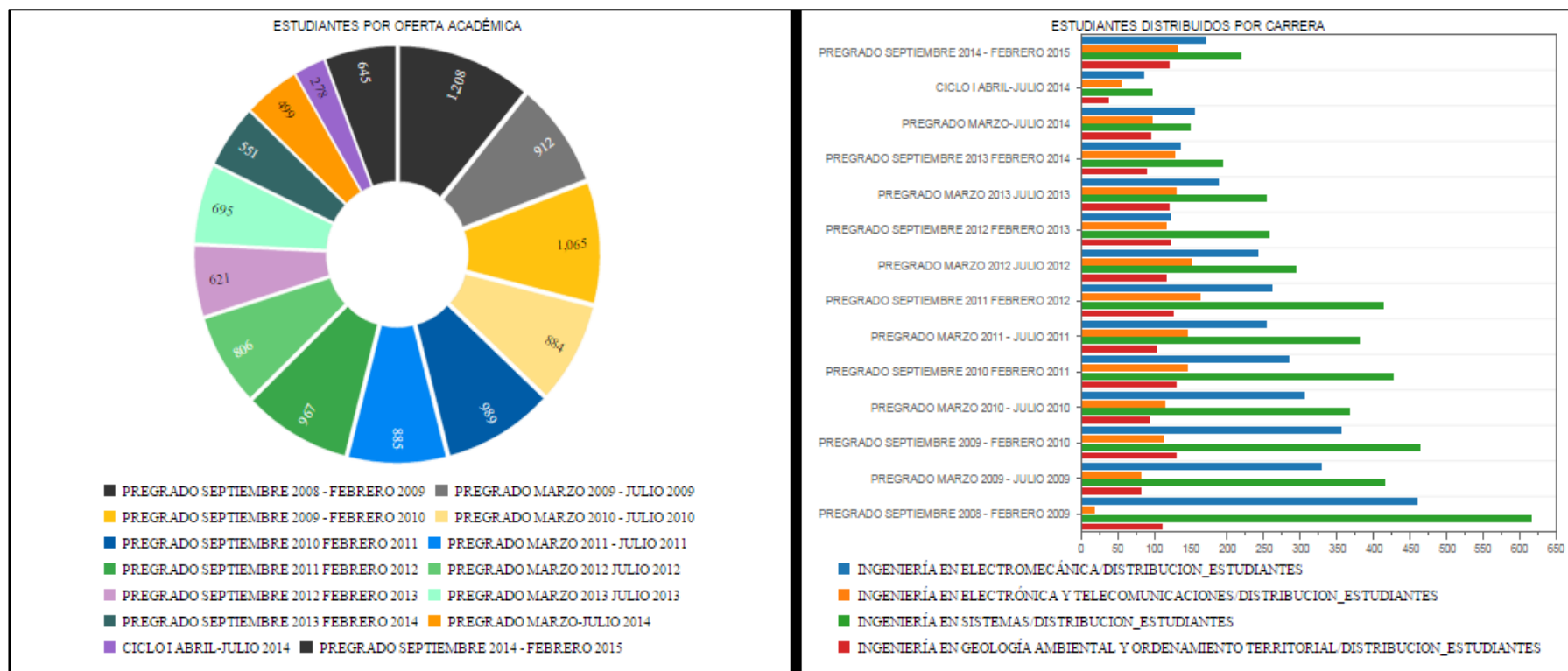


Figura 41 Tasa de estudiantes por Oferta Académica – Distribución estudiantes por Carrera.



G. DISCUSIÓN.

1. DESARROLLO DE LA PROPUESTA ALTERNATIVA.

La propuesta alternativa describe el proceso realizado para el cumplimiento de los objetivos.

- ✓ **Recopilar información para el entendimiento del negocio o departamento pertinente.**

Dentro de este objetivo, en primera instancia se realizó una búsqueda y recopilación de información acerca del AEIRNNR, mediante entrevista realizada al director de la misma, esta entrevista permitió tener una visión más clara del negocio, es decir conocer la naturaleza, los procesos que se llevan a cabo dentro del área, las necesidades de información que existen en cuanto a la parte académica se refiere, las fuentes de datos de las que dispone, los sistemas transaccionales, etc.

Adicional a la técnica de la entrevista, se aplicó la observación directa como instrumento de apoyo, la combinación de ambas técnicas nos permitieron de una u otra manera encontrar falencias en la parte de la gestión de la información académica del AEIRNNR, ya que no se dispone de un sistema dedicado a dicha actividad. Otro factor a considerar, es el desconocimiento total que existe acerca del término Inteligencia de Negocios, de allí la necesidad de divulgar mucho más este tipo de tecnologías, mediante la implantación de sistemas que estén encaminados a este tipo de actividades, y que aporten eficientemente en el proceso de toma de decisiones de cualquier organización.

- ✓ **Investigar y seleccionar la metodología para la implementación de Inteligencia de Negocios.**

Para el cumplimiento del presente objetivo, se indagó acerca de las diferentes tipos de metodologías que existen acerca de Inteligencia de Negocios, si bien es cierto existe una gama bastante considerable de metodologías enfocadas a este tipo de soluciones, por tal motivo se realizó un análisis de cada una de ellas, y se determinó que las más sobresalientes son la Metodología Inmon [9] y Kimball [8], ya que han sido desarrolladas por los que son considerados como los padres del Data Warehousing. Adicional a estas metodologías, se investigó acerca de la metodología HEFESTO, que es un tipo de metodología híbrida que puede adaptarse a cualquiera de las dos antes mencionadas, con la diferencia de que esta última, dispone de bastante



documentación en español, siendo esta la que mejor se adapta para la realización del proyecto.

✓ **Implantación del modelo de componentes de Inteligencia de Negocios en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables.**

Para el desarrollo y cumplimiento del objetivo, se tomó como base toda la información recolectada acerca del AEIRNNR, y siguiendo cada uno de los pasos de la metodología HEFESTO se empezó a identificar los indicadores y perspectivas de análisis, para luego modelar cada uno de los Data Marts (estudiantes, docentes) que representarán la información académica del AEIRNNR. Posteriormente se construyó cada uno de los procesos ETL que nos permitieron extraer los datos necesarios para poblar cada uno de los Data Marts, y así poder responder a las necesidades de información académica que existen dentro del AEIRNNR,

Los procesos ETL se los llevo a cabo mediante la herramienta Pentaho Data Integration, la misma que nos proporcionó cada una de los componentes necesarios para extraer la información del SGASWebServices. La información inicialmente fue extraída desde el SGASWebServices a una base de datos temporal, para posteriormente, mediante técnicas de caracterización de datos asegurar la calidad de los mismos.

Una vez realizada la caracterización de datos, se poblaron cada uno de los Data Marts, todo este proceso fue guiado mediante la metodología HEFESTO. Los procesos ETL, dentro de cualquier proyecto Data Warehousing representan el 80 % de la totalidad del mismo, ya que estos conllevan un alto grado de complejidad, más aún cuando los OLTP de los que se dispone no se encuentran dentro de un esquema o base de datos relacional. Para la presentación de información al usuario final se generó reportes (Pentaho Report Designer), cubos OLAP (Pentaho Schema Workbench), los mismos que podrán ser visualizados mediante el módulo Pentaho BI Server, que mediante ejecución local en el puerto 8080 nos proporcionó una interfaz web, para navegar por los diferentes reportes desarrollados con los módulos mencionados anteriormente.



2. VALORACIÓN TÉCNICA ECONÓMICA AMBIENTAL.

El Trabajo de Titulación denominado “**Componentes de Inteligencia de Negocios en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables**”, desde el punto de vista técnico, se lo puede considerar como un trabajo factible y necesario dentro de la dirección del área, puesto que servirá como apoyo para futuras investigaciones académicas, exclusivamente de la carrera de Ingeniería en Sistemas, ya que se puede aplicar técnicas de minería de datos, en base a la información histórica contenida en el Data Warehouse.

Desde el punto de vista económico, el desarrollo del proyecto se ajusta a los intereses de la institución, así como a las necesidades de información que existen dentro de la misma, y el hecho de trabajar con herramientas Open Source, implica el no tener que pagar por los costos de licencias del software.

El trabajo de titulación fue desarrollado en su totalidad por el investigador, teniendo de guía a un docente de la institución para la elaboración del anteproyecto de tesis, y un segundo docente que será el director del trabajo de titulación. La Tabla XV (*Recursos Humanos*) detalla el tiempo y costo asignado al investigador y docente director del proyecto.

TABLA XV RECURSOS HUMANOS.

Descripción	Cantidad	N. Horas	V. Unitario	Total
Investigador	1	400	\$ 5,00	\$ 2000,00
Director	1	100	\$ 5,00	\$ 500,00
TOTAL				\$ 2500,00

La tabla XVI (*Servicios Básicos*) muestra los servicios básicos que se utilizó para el desarrollo del proyecto, los mismos que serán cubiertos por el investigador.

TABLA XVI SERVICIOS BÁSICOS.

Descripción	Cantidad	V/unitario	Total
Transporte	80	\$ 0,25	\$ 20,00
Internet	1	\$ 20,00	\$ 20,00
SW libre	1	\$ -	\$ -
TOTAL:			\$ 40,00



La Tabla XVII (*Recursos Materiales*) presenta una descripción detallada de los recursos materiales que fueron necesarios para la presentación de los avances y el informe final del proyecto.

TABLA XVII RECURSOS MATERIALES.

M. Oficina	Cantidad	V/Unitario	Total
Cartucho de tinta	8	\$ 5,00	\$ 40,00
Resma de papel	4	\$ 3,50	\$ 14,00
Anillados	6	\$ 1,50	\$ 9,00
Perfiles	2	\$ 0,70	\$ 1,40
Cds	4	\$ 1,00	\$ 4,00
Copias	360	\$ 0,02	\$ 7.20
Total			\$ 75.60

La Tabla XVIII (*Recursos Hardware*) detalla los recursos hardware utilizados en el desarrollo del proyecto, la misma que comprende un computador portátil usado para la implementación de los Data Marts y la redacción de los avances e informe final del mismo.

TABLA XVIII RECURSOS HARDWARE.

Equipo	V. Real	T. Vida (años)	V. Dep/año	Total
Laptop Toshiba	\$ 1100	3	\$ 366.66	\$ 366.66
Impresora	\$ 50	1	\$ 4.16	\$ 4.16
Flash memory 4 GB Kingston	\$ 10	3	\$ 3.33	\$ 3.33
TOTAL:				\$ 374.15

La Tabla XIX (*Recursos Software*) muestra el software utilizado para la realización del proyecto, se debe mencionar que todas las herramientas utilizadas son open source, lo que implica el no tener que pagar por la adquisición de licencias.



TABLA XIX RECURSOS SOFTWARE.

Descripción	Valor
PostgreSQL	\$ 00,00
Pentaho Data Integration	\$ 00,00
Pentaho Report Designer	\$ 00,00
Pentaho Schema Worbench	\$ 00,00
Community Dashboard Editor	\$ 00,00
Pentaho BIServer	\$ 00,00
PgAdmin III	\$ 00,00
JAVA™ SE 7	\$ 00,00
SQL Power Architect	\$ 00,00
TOTAL:	\$ 00,00

En la Tabla XX (*Recursos Técnicos y Tecnológicos*) se aprecia la suma parcial de los recursos técnicos y tecnológicos.

TABLA XX RECURSOS TÉCNICOS Y TECNOLÓGICOS.

Descripción	Valor Total
Recursos Hardware	\$ 374.15
Recursos Software	\$ 00.00
TOTAL:	\$ 374.15

La Tabla XXI (*Costo del Proyecto (aproximación)*), nos muestra la suma total de todos los recursos: humanos, materiales, técnicos y tecnológicos usados en el presente proyecto, lo que nos brinda una aproximación real del coste del proyecto.

TABLA XXI COSTO DEL PROYECTO (APROXIMACIÓN).

Gastos	Total
Talento Humano	\$ 2500,00
Recursos Técnicos y Tecnológicos	\$ 374,15
Recursos Físicos	\$ 75,60
Servicios Básicos	\$ 40,00
Subtotal	\$ 2989,75
Imprevistos	\$ 298,975
Total	\$ 3288,725



Desde el punto de vista ambiental el proyecto sería de gran aporte para el medio ambiente, ya que el hecho de generar reportes de manera digital, reduciría en un cierto porcentaje el uso de papel, por lo que el impacto que genera el proyecto es mínimo.



H. CONCLUSIONES.

- La inteligencia de negocios mediante actividades analíticas, permite a una organización conocer su pasado, presente y futuro, sirviendo de apoyo en el proceso de toma de decisiones a nivel estratégico, táctico y operativo de la organización
- Es fundamental seleccionar y evaluar una arquitectura de inteligencia de negocios que se ajuste al presupuesto y recursos de la organización, ya que si bien es cierto existen dos arquitecturas de inteligencia de negocios (Ralph Kimball, Bill Inmon), pero se debe considerar que ambas poseen variantes significativas, que pueden afectar el presupuesto y los recursos de un proyecto.
- Las herramientas Open Source, permiten dar soluciones de inteligencia de negocios a las organizaciones, tal es el caso de Pentaho Business Intelligence, que nos proporciona la suite completa de inteligencia de negocios, lo que implica tener todo el sistema Data Warehouse integrado bajo una misma plataforma.
- La arquitectura de flujo de datos es de vital importancia al momento de emprender un proyecto de inteligencia de negocios en cualquier organización, ya que nos permite apreciar los esquemas donde se depositarán los datos durante todo el proceso ETL, además las actividades que se realizarán durante cada etapa del flujo de carga de datos.
- La caracterización de datos es fundamental en un proceso ETL, ya que esta nos permite presentar al usuario final información limpia y de calidad.
- El proceso de limpieza y transformación de los datos provenientes de las diferentes fuentes de información, es un proceso extenso e importante, que afectará profundamente en los resultados que se desea obtener.



I. RECOMENDACIONES.

- Realizar un plan de mantenimiento y actualización del Sistema de Gestión Académico y del SGASWebServices, debido a que este último contiene demasiada información inconsistente o errónea.
- Utilizar una metodología de desarrollo iterativa para ambientes de producción, ya que un negocio es muy variable y el sistema Data Warehouse debe ser lo suficientemente flexible, para responder a preguntas del mismo conforme este va evolucionando.
- Utilizar el Data Warehouse para ampliar el alcance del proyecto, mediante técnicas de minería de datos en base a la información histórica contenida en el mismo, podría resultar interesante en un proyecto de titulación para futuros ingenieros.
- Utilizar herramientas Open Source, debido a que estas poseen la misma funcionalidad que cualquier herramienta privativa, lo que implica tener acceso al código de las mismas y el no tener que pagar por algún tipo de licencia.
- Disponer de un plan de internet de alta velocidad, con el fin de mejorar el rendimiento de los procesos ETL, y evitar inconvenientes al momento de realizar los procesos de actualización del Data Warehouse.



J. BIBLIOGRAFÍA.

- [1]. Salvador Grandulfo. (Agosto, 2010). Inteligencia de Negocios. Disponible en: <http://www.emb.cl/gerencia/articulo.mvc?xid=219>. Último acceso: 15 de octubre de 2014.
- [2]. Definición de Business Intelligence. Disponible en: <http://www.ibermatica.com/ibermatica/businessintelligence2>. Último acceso: 15 de octubre de 2014.
- [3]. Bernabéu Ricardo Darío. Data Warehousing Investigación y Sistematización de conceptos. Disponible en: http://www.linuxbi.com/files/datawarehouse_hefesto.pdf. Último acceso: 15 de octubre de 2014.
- [4]. Hugo Sans. Inteligencia de Negocios con Pentaho. Disponible en: <http://pentahoparatodos.blogspot.com/2014/03/datawarehouse-almacenes-de-datos.html>. Último acceso: 15 de octubre de 2014.
- [5]. Sinnexus. Bussines Intelligence. Informática estratégica. Disponible en: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx
- [6]. Dataprix. ¿Qué es datawarehouse? Disponible en: <http://www.dataprix.com/qu-es-un-data-warehouse>. Último acceso: 15 de octubre de 2014.
- [7]. Inteligencia y Competitividad. Inteligencia de Negocios. Disponible en: <http://journals.continental.edu.pe/index.php/ParaEmprender/article/view/164/163>. Último acceso: 15 de octubre de 2014.
- [8]. R. Kimball – Margy Ross. The Data Warehouse Toolkit. Tercera Edicion. John Wiley & Sons, Inc. 2013.
- [9]. IMHOFF, Claudia; GALEMMO, Nicholas; GEIGER, Jonathan G. Mastering Data Warehouse Design. Primera Edicion. Wiley publishing, Inc. 2003.
- [10]. LAUDON, Kennetth. C; LAUDON, Jane. P. Sistemas de Información Gerencial. Décima Edición. Pearson Education, Inc. 2003.
- [11]. Gartner. Cuadrante mágico para plataformas de Inteligencia de Negocios. Disponible en: <http://www.gartner.com/technology/reprints.do?id=1-1QLGACN&ct=140210&st=sb>. Último acceso: 16 de octubre del 2014.
- [12]. Pentaho. Soluciones de Inteligencia de Negocios. Disponible en: <http://www.pentaho.com/>. Último acceso: 16 de octubre del 2014.



- [13]. Dataprix. Metodología Hefesto. Disponible en: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/hefesto-metodologia-propia-para-la-construccion-un-data-wareh>. Último acceso: 16 de octubre del 2014.
- [14]. C. M. Murillo. "Implementación de un Data Mart para el ramo de Vehículos en una empresa aseguradora". Trabajo de fin de carrera. Facultad de ingeniería de Sistemas. Escuela Politécnica Nacional. Quito, 2008. Disponible en: <http://bibdigital.epn.edu.ec/bitstream/15000/1281/1/CD-2034.pdf>
- [15]. C. P. López - D. S. González. Minería de datos, Técnicas y Herramientas. Primera Edición. Thomson Ediciones Paraninfo, S.A, 2007. Disponible en: http://books.google.es/books?hl=es&lr=&id=wz-D_8uPFCEC&oi=fnd&pg=PR4&dq=mineria+de+datos&ots=ThU1yh4x6l&sig=ZbDpjpGt-wAox8O4Anu4mtj7v0E#v=onepage&q=mineria%20de%20datos&f=false
- [16]. Stratebi. Open Business Intelligence Spago BI. Disponible en: <http://www.stratebi.com/spagobi>. Último acceso: Último acceso: 28 de octubre del 2014.
- [17]. Talend Open Studio. Disponible en: <https://www.talend.com/products/talend-open-studio>. Último acceso: Último acceso: 28 de octubre del 2014. Último acceso: 28 de octubre del 2014.
- [18]. J. Curto Díaz - J. Conesa. Introducción al Business Intelligence. Primera Edición. Editorial El Ciervo 96, SA, Mayo 2010. [En Línea]. Disponible en: <http://books.google.com.ec/books?id=iU3RAXYQXMkC&pg=PA37&lpg=PA37&dq=arquitectura+de+un+data+warehouse&source=bl&ots=Nn8enp64qr&sig=HPwy6fDjl8q5VsrnYM4AP199ipw&hl=es&sa=X&ei=D01RVLWGMsSeNqPKgugG&ved=0CD0Q6AEwBDgU#v=onepage&q=arquitectura%20de%20un%20data%20warehouse&f=false>
- [19]. A. Peña Ayala. Inteligencia de Negocios: Una propuesta para su desarrollo en las Organizaciones. Primera Edición. Instituto Politécnico Nacional, México, 2006. Disponible en: http://www.wolnm.org/apa/articulos/Inteligencia_Negocios.pdf
- [20]. Phil Bartle. Información para la gestión y gestión de la información. Disponible en: <http://cec.vcn.bc.ca/mpfc/modules/mon-miss.htm>. Último acceso: 28 de octubre del 2014.



- [21]. D. Arenas C. Análisis BI Tecnologías, Metodologías, Opiniones, Eventos, Herramientas, Evaluaciones, Recomendaciones, Noticias, Modelos, Mejores Prácticas, Análisis, y mucho más relacionado con el mundo de Business Intelligence. Disponible en: <http:// analisisbi.blogspot.com/2009/10/sistemas-etl.html>. Último acceso: 28 de octubre del 2014.
- [22]. P. Urquizu. Business Intelligence fácil. Disponible en: <http://www.businessintelligence.info/serie-dwh/claves-subrogadas.html>. Último acceso: 05 de noviembre del 2014.
- [23]. M. Murillo. Data Warehousing y Business Intelligence con Herramientas Open Source. Disponible en: <http://ingmmurillo-dwh-bi.blogspot.com/2012/06/arquitectura-de-flujos-de-datos-con.html>. Último acceso: 05 de noviembre del 2014.
- [24]. M. Golfarelli. Open Source BI Platforms: a Functional and Architectural Comparison. Disponible en: <http://bias.csr.unibo.it/golfarelli/papers/DAWAK09%20-%20Golfarelli.pdf>. Último acceso: 05 de noviembre del 2014.
- [25]. ECMA-404. The JSON Data Interchange Standard. Introducción a Json. Disponible en: <http://www.json.org/json-es.html>. Último acceso: 05 de noviembre del 2014.
- [26]. J.C. Trujillo, J. N. Mazón, J. Pardillo. Diseño y explotación de almacenes de datos. Conceptos Básicos de Modelado Multidimensional. Primera Edición. Club Universitario 2010. Disponible en: <http://books.google.com.ec/books?id=E7Aceg-o4oC&pg=PA36&lpg=PA36&dq=dimensiones+degeneradas&source=bl&ots=nm0Uq9LNqe&sig=Pv11Acly-Tf7NRib7Q4ZY-EwHd4&hl=es&sa=X&ei=EUB3VNG8I4KmNtaYgNAH&ved=0CFkQ6AEwCQ#v=onepage&q=dimensiones%20degeneradas&f=false>. Último acceso: 27 noviembre 2014.
- [27]. A. Silberschatz. Fundamentos de bases de datos. Cuarta edición. Editorial McGRAW-HILL.
- [28]. R. Elmasri, S. Navathe. Fundamentos de sistemas de Bases de Datos. Quinta edición. Editorial PEARSON EDUCACIÓN S.A, Madrid, 2007.
- [29]. Anthony J. Gómez Morales. (2012) Inteligencia de negocios, una ventaja competitiva para las organizaciones. [En Línea]. Disponible en: <http://revistas.unitru.edu.pe/index.php/PGM/article/view/193/199>



- [30]. R. Espinosa. (Julio, 2010). Inteligencia de Negocios con Pentaho. [En Línea]. Disponible en: <https://churriwifi.wordpress.com/category/pentaho/>
- [31]. M. Tamayo. (Diciembre 2006). Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP. [En Línea]. Disponible en: <http://www.redalyc.org/articulo.oa?id=64326317>
- [32]. M. A. Ibarra. (2006). Procesamiento Analítico en Línea (OLAP). [En Línea]. Disponible en: <http://exa.unne.edu.ar/informatica/SO/OLAPMonog.pdf>



K. ANEXOS.

Anexo 1: Anteproyecto del Trabajo de Titulación.



**UNIVERSIDAD
NACIONAL
DE LOJA**

PFC-CIS-UNL



Área de la Energía, las Industrias y los Recursos Naturales No Renovables

CARRERA DE INGENIERÍA EN SISTEMAS

**“Componentes de Inteligencia de Negocios en
el Área de la Energía, las Industrias y los
Recursos Naturales no Renovables”.**

PROYECTO FIN DE CARRERA.

Autor: Tuza – Cuenca, Edgar – Iván.

Tutor: Ing. Paz – Arias, Henry - Patricio.

Loja – Ecuador
2014.



Índice

A.	Tema.....	82
B.	Problemática.....	83
C.	Justificación.....	85
D.	Objetivos.....	87
E.	Alcance.....	88
F.	Metodología.....	89
G.	Cronograma.....	90
H.	Presupuesto.....	91
I.	Bibliografía.....	93
J.	Anexos.....	94



Índice de tablas.

TABLA XXII TALENTO HUMANO PARA EL DESARROLLO DEL PROYECTO.....	91
TABLA XXIII RECURSOS TÉCNICOS PARA EL DESARROLLO DE PROYECTO....	91
TABLA XXIV SERVICIOS BÁSICOS PARA EL DESARROLLO DE PROYECTO.....	91
TABLA XXV RECURSOS FÍSICOS PARA EL DESARROLLO DE PROYECTO.	92
TABLA XXVI PRESUPUESTO GENERAL PARA EL DESARROLLO DEL PROYECTO.	92



A. Tema.

“Componentes de Inteligencia de Negocios en el Área de la Energía, las Industrias y los Recursos Naturales No Renovables”.



B. Problemática.

1. Situación Problemática.

Hasta hace poco la inteligencia de negocios parecían reservadas para empresas del sector comercial, y de hecho este tipo de herramientas están muy poco difundidas entre el segmento de organizaciones educativas de nivel superior [2]. Esta situación viene provocada por el elevado coste de las licencias de estas soluciones informáticas, así como los prolongados periodos de desarrollo e implantación de las mismas [1].

Sin embargo, muchos directores de Universidades están comprendiendo la importancia de la gestión de la información y las ventajas competitivas que implica su uso [9]. Este proceso de gestión consiste en resolver de una manera eficiente los problemas que existen en el análisis de distintos tipos de datos de la organización y su entorno, a través de la explotación de la información por medio de las tecnologías de la información (TI), facilitando la adaptación de aplicaciones para la inteligencia de negocios [8].

Con los sistemas tradicionales se preparan reportes para encontrar las respuestas a algunas preguntas [3], pero se necesita dedicar demasiado tiempo asignado al análisis de localización y presentación de los datos, como también asignación de recursos humanos y de procesamiento del departamento de sistemas para poder responderlas, sin tener en cuenta la degradación de los sistemas transaccionales [4].

Esta problemática se debe a que dichos sistemas transaccionales no fueron construidos con el fin de brindar síntesis, análisis, consolidación, búsquedas y proyecciones [8]. Sin embargo, pese a no tener el soporte financiero suficiente como para adquirir herramientas o soluciones costosas, existe un desconocimiento total sobre herramientas open source [5] por parte de los directivos, lo que acarrea consigo un sinnúmero de problemas como la resistencia al cambio, y la falta de predisposición para cubrir necesidades de integrar y depurar toda la información mediante la implantación de un Data Warehouse, lo que implica que no se cuente con reportes dinámicos, ágiles y de gran capacidad que permitan realizar un estudio detallado de la información, pero sobre todo, el no poder compartir y distribuir el conocimiento de la organización provocando ineficiencia en la toma de decisiones [3].



2. Problema de Investigación.

Tomando en cuenta los problemas mencionados anteriormente, se puede constatar que el principal problema que existe dentro del Área de la Energía, las Industrias y los Recursos Naturales no Renovables de la Universidad Nacional de Loja es:

¿La ineficiencia de los sistemas de información operacionales, satisfacen los requerimientos de información, y causan impacto en la toma de decisiones?



C. Justificación.

La necesidad que tienen las Instituciones de Educación Superior de ajustarse rápidamente a los requerimientos de información, y a los cambios en el ambiente económico y tecnológico, hace necesario que la administración requiera de información oportuna y actualizada [9].

Por ello, es pertinente implementar Inteligencia de Negocios dentro del Área de la Energía, las Industrias y los Recursos Naturales no Renovables de la Universidad Nacional de Loja, para contribuir a mejorar el proceso de toma de decisiones en el área académica, elevar la ventaja competitiva, y contar con una mejor planificación de los recursos; además, generar mayor eficiencia operativa, obtener una mayor comprensión y entendimiento del entorno y de los procesos desde la propia experiencia en las personas, un acceso rápido y fácil a la información, y un entorno gráfico de los datos que contribuya a mejorar el proceso de toma de decisiones en el área administrativa, obtener una vista precisa del rendimiento de la institución mediante el uso de métricas, conocer el comportamiento y preferencias de estudiantes y profesores, identificar los patrones y tendencias dentro de grandes volúmenes de datos [8].

La implementación de inteligencia de negocios, como parte de una solución proporcionará información útil a nivel estratégico y administrativo, sirviendo de apoyo en la toma de decisiones [4], para ello se dispone de herramientas open source, que nos permiten centralizar toda la información y generar reportes personalizados y dinámicos [6], sirviendo de apoyo en la toma de decisiones, produciendo más eficiencia, agilidad en los procesos actuales, mayor velocidad de comunicación, información en tiempo real, lo que significa eficiencia y eficacia en la toma de decisiones.

Desde el punto de vista económico, la inteligencia de negocios cambia los costos relativos del capital, así como los costos de la información de la empresa [7], y el hecho de trabajar con herramientas Open Source, implica el no tener que pagar por los costos de licencias del software [5]. Evidentemente esta tecnología tiene su impacto social, ya que el hecho de implementarla, provoca una reducción en el número de gerentes intermedios y oficinistas [7], pero desde otra perspectiva sería de gran aporte para el medio ambiente, ya que el hecho de generar reportes de manera digital, reduciría en un cierto porcentaje el uso de papel. Por lo antes expuesto la



realización del presente proyecto se justifica plenamente desde el punto de vista académico, técnico-tecnológico, económico, y ambiental.



D. Objetivos.

1. Objetivo General.

Incorporar Componentes de Inteligencia de Negocios en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables de la Universidad Nacional de Loja.

2. Objetivos Específicos.

- Recopilar información para el entendimiento del negocio o departamento pertinente.
- Investigar y seleccionar la metodología para la implementación de inteligencia de Negocios.
- Implantación del modelo de componentes de inteligencia de negocios en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables de la Universidad Nacional de Loja.



E. Alcance.

La necesidad de implantar componentes de herramientas de Inteligencia de Negocios dentro de la Universidad Nacional de Loja, surge debido a que los sistemas actuales operacionales no satisfacen las necesidades de información que apoyen a los administrativos en la toma de decisiones, por lo que se plantea desarrollar el presente proyecto, para así analizar el impacto que genera la aplicación de Inteligencia de Negocios en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables de la Universidad Nacional de Loja, el mismo que constara de las siguientes fases:

1. Recopilación de información.

- Analizar y comprender la naturaleza del negocio.
- Obtener los requerimientos e indicadores claves del negocio.
- Analizar las distintas fuentes de datos del negocio.

2. Selección de la metodología.

- Seleccionar la metodología para la implementación de inteligencia de negocios.

3. Fase de Implantación.

- Modelo lógico y físico del Data Warehouse.
 - Tablas de dimensiones.
 - Tablas de hechos.
 - Procesos ETL.
 - Verificación y refinamiento del modelo lógico para determinar su consistencia.
- Extracción de la información.
 - Generar reportes dinámicos.
 - Generar reportes personalizados.



F. Metodología.

Para el desarrollo del presente trabajo se utilizarán los siguientes métodos de investigación:

- **Método Científico.**

Este método nos ayudará a durante toda la ejecución del proyecto, ya que nos proporcionará el camino a seguir desde la observación del problema hasta las conclusiones, mediante la obtención y análisis de los factores que influyen en la implementación de la inteligencia de negocios.

- **Método Analítico.**

Este método nos ayudará a observar y determinar la naturaleza del objeto en estudio, para así poder determinar las causas y los efectos del mismo. Además este método nos permitirá explicar y comprender un poco más a fondo sobre el funcionamiento en sí de la inteligencia de negocios.

- **Método Inductivo.**

Este método se lo utilizará para estructurar la problemática y justificación, buscando información relacionada con el tema y permitiéndonos dar solución a los diferentes problemas que se presentan en las empresas cuando no poseen herramientas de inteligencia de negocios.

- **Metodología para Inteligencia de negocios.**

La metodología para la implementación inteligencia de negocios nos ayudara a llevar un proceso más ordenado y secuencial, sobre las diferentes etapas del proceso de implementación de inteligencia de negocios.

Técnicas de recolección de información:

- **Observación.**

Esta técnica se la utilizara para tener una observación de campo, para constatar la problemática que ocurre dentro del área en cuanto a la gestión de información académica.

- **Entrevista.**

Se la utilizara para obtener mayor información acerca de los requerimientos de información que tienen dentro del área, así como las limitaciones que tienen los sistemas de información que se manejan dentro de la misma.

G. Cronograma.

	Modo de tarea ▼	Nombre de tarea ▼	Duración ▼	Trabajo ▼	Comienzo ▼	Fin ▼
1	★	▸ Recopilación de información.	10 días	20 horas	mar 13/05/14	mié 28/05/14
2	★	Analizar y comprender la naturaleza del negocio.	2 días	4 horas	mar 13/05/14	mié 14/05/14
3	★	Obtener los requerimientos e indicadores claves del negocio.	2 días	4 horas	jue 15/05/14	vie 16/05/14
4	★	Analizar las distintas fuentes de datos del negocio.	6 días	12 horas	mar 20/05/14	mié 28/05/14
5	★	▸ Selección de la metodología.	6 días	12 horas	jue 29/05/14	vie 06/06/14
6	★	Seleccionar la metodología para la implementación de inteligencia de negocios.	6 días	12 horas	jue 29/05/14	vie 06/06/14
7	★	▸ Fase de Implantación.	184 días	368 horas	mar 10/06/14	vie 24/04/15
8	★	Tablas de dimensiones.	38 días	76 horas	mar 10/06/14	mié 13/08/14
9	★	Tablas de hechos.	38 días	76 horas	jue 14/08/14	vie 17/10/14
10	★	Procesos ETL.	38 días	76 horas	mar 21/10/14	mié 24/12/14
11	★	Verificación y refinamiento del modelo lógico para determinar su consistencia	38 días	76 horas	jue 25/12/14	vie 27/02/15
12	★	Generar reportes dinámicos.	16 días	32 horas	mar 03/03/15	vie 27/03/15
13	★	Generar reportes personalizados.	16 días	32 horas	mar 31/03/15	vie 24/04/15

Figura 42 Cronograma del Proyecto de Titulación.



H. Presupuesto.

1. Talento Humano:

TABLA XXII TALENTO HUMANO PARA EL DESARROLLO DEL PROYECTO.

Descripción	Cantidad	N. Horas	V. Unitario	Total
Investigador	1	400	\$ 5,00	\$ 2000,00
Director	1	100	\$ 5,00	\$ 500,00
TOTAL				\$ 2500,00

2. Recursos técnicos:

TABLA XXIII RECURSOS TÉCNICOS PARA EL DESARROLLO DE PROYECTO.

Equipo	V. Real	T. Vida (años)	V. Dep/año	Total
Laptop Toshiba	\$ 1100	3	\$ 366,66	\$ 366,66
Impresora	\$ 50	1	\$ 4,16	\$ 4,16
Flash memory 4 GB Kingston	\$ 10	3	\$ 3,33	\$ 3,33
TOTAL:				\$ 374,15

3. Recursos básicos:

TABLA XXIV SERVICIOS BÁSICOS PARA EL DESARROLLO DE PROYECTO.

Descripción	Cantidad	V/unitario	Total
Transporte	80	\$ 0,25	\$ 20,00
Internet	1	\$ 20,00	\$ 20,00
SW libre	1	\$ -	\$ -
TOTAL:			\$ 40,00



4. Recursos físicos:

TABLA XXV RECURSOS FÍSICOS PARA EL DESARROLLO DE PROYECTO.

M. Oficina	Cantidad	V/Unitario	Total
Cartucho de tinta	8	\$ 5,00	\$ 40,00
Resma de papel	4	\$ 3,50	\$ 14,00
Anillados	6	\$ 1,50	\$ 9,00
Perfiles	2	\$ 0,70	\$ 1,40
Cds	4	\$ 1,00	\$ 4,00
Copias	360	\$ 0,02	\$ 7.20
Total			\$ 75,60

5. Total de recursos:

TABLA XXVI PRESUPUESTO GENERAL PARA EL DESARROLLO DEL PROYECTO.

Gastos	Total
Talento Humano	\$ 2500,00
Recursos Técnicos	\$ 374,15
Recursos Físicos	\$ 75,60
Servicios Básicos	\$ 40,00
Subtotal	\$ 2989,75
Imprevistos	\$ 298,975
Total	\$ 3288,725

6. Financiamiento:

Debido a que es un proyecto de carácter académico los gastos serán cubiertos por el tesista.



I. Bibliografía.

- [1]. Ignacio Maroto. (Agosto, 2010). Inteligencia de negocios. [En línea]. Disponible en: <http://emprendeme.blogspot.com/2010/08/seguramente-muchos-de-vosotros-habreis.html>
- [2]. Antonio Muñoz Cañavate. (Mayo, 2003). Sistemas de información en las empresas. Disponible en: http://www.upf.edu/hipertextnet/numero-1/sistem_infor.html
- [3]. Juan Carlos Aranibar S. (Mayo, 2003). Inteligencia de negocios. Disponible en: http://www.scielo.org.bo/scielo.php?pid=S207733232003000100010&script=sci_arttext
- [4]. Tendencias de la Inteligencia de negocios. (Agosto, 2007). [En Línea] Disponible en: <http://investiga.ide.edu.ec/index.php/revista-agosto-2007/703-las-nuevas-tendencias-de-la-inteligencia-de-negocios>
- [5]. Anthony J. Gómez Morales. (2012) Inteligencia de negocios, una ventaja competitiva para las organizaciones. [En Línea]. Disponible en: <http://revistas.unitru.edu.pe/index.php/PGM/article/view/193/199>
- [6]. Roberto Espinosa. (Julio, 2010). Inteligencia de Negocios con Pentaho. [En Línea]. Disponible en: <https://churriwifi.wordpress.com/category/pentaho/>
- [7]. Kenneth C. Laudon, Jane P. Laudon. Sistemas de Información Gerencial. Décima Edición. PEARSON Education, México 2008.
- [8]. Javier Nader. Sistema de Apoyo Gerencial Universitario. [En Línea]. Disponible en: <http://www2.itba.edu.ar/archivos/secciones/nader-tesisdemagister.pdf>
- [9]. Luis Fuentes Tapia, Ricardo Valdivia Pinto. Incorporación de elementos de inteligencia de negocios en el proceso de admisión y matrícula de una universidad chilena. [En Línea]. Disponible en: http://www.scielo.cl/scielo.php?pid=S071833052010000300012&script=sci_arttext&tlng=en
- [10]. Paula Andrea Castaño Cárdenas. Estudio de la Evolución de los sistemas Open Source de BI durante la última década. [En Línea]. Disponible en: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/8178/1/pcastanocTFM0611.pdf>



J. Anexos.

1. LICENCIA CREATIVE COMMONS.



Componentes de Inteligencia de Negocios en el Área de la Energía, las Industrias y los Recursos Naturales no Renovables por Edgar Iván Tuza Cuenca

se distribuye bajo una Licencia [Creative Commons](#)

[Atribución-NoComercial 4.0 Internacional.](#)



Anexo 2: Modelo de Entrevista.

MODELO DE ENTREVISTA PARA LA RECOLECCIÓN DE INFORMACIÓN.

- **Entrevistado:** Ing. Jorge Gahona Pacheco. Director del Área de la Energía, las Industrias y los Recursos Naturales No Renovables.
- **Entrevistador:** Edgar Iván Tuza Cuenca. Egresado de la carrera de Ingeniería en Sistemas del Área de la Energía, las Industrias y los Recursos Naturales No Renovables de la Universidad Nacional de Loja.

Preguntas:

1. ¿Conoce Ud. acerca de Inteligencia de Negocios?
2. ¿El AEIRNNR cuenta actualmente con un sistema dedicado a generar reportes con la información académica de la misma?
3. ¿Dedica Ud. Horas extras al análisis de documentos e informes relacionados con la parte académica del AEIRNNR?
4. ¿Ha perdido Ud. oportunidades de negocio por la falta de disponibilidad de información?
5. ¿Considera factible o necesaria la instalación de componentes de inteligencia de negocios dentro de la dirección del AEIRNNR?
6. ¿Cuál es el principal sistema informático que procesa la información académica del AEIRNNR?
7. ¿Con qué frecuencia se realizan transacciones en el sistema del AEIRNNR?
8. ¿A qué cantidad de información académica Ud. tiene acceso?
9. En cuanto a la información académica del AEIRNNR, ¿en qué tipo de archivos se almacena esta, quien es la persona o departamento encargado?
10. ¿La información académica del AEIRNNR se encuentra centralizada o distribuida en diferentes repositorios?
11. ¿Cuáles son los indicadores claves de rendimiento, que le permiten a Ud. medir el desenvolvimiento académico del AEIRNNR?
12. ¿Cuáles son las perspectivas de análisis, por las cuales desea analizar la información académica del AEIRNNR?
13. Si se llegara a implementar un Data Warehouse dentro del AEIRNNR, ¿Cuál considera Ud. que debería ser el período de actualización?
14. En cuanto a la entrega de información y reportes, ¿en qué formato le gustaría recibir la información?



15. En cuanto a la información histórica del AEIRNNR, ¿a partir de qué año o período le gustaría disponer de dicha información?

CERTIFICACIÓN:

Ing. Jorge Gahona Pacheco
1101061172

**DIRECTOR DEL ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS
NATURALES NO RENOVABLES.**



Anexo 3: Diccionario de datos.

1. Dimensión DIM_CARRERA.

La dimensión DIM_CARRERA, representa cada una de las carreras que se ofrecen dentro del AEIRNNR, esta dimensión nos permitirá realizar un análisis del rendimiento académico de la población estudiantil por carreras, además el nivel de preparación de los docentes, etc.

TABLA XXVII DIMENSIÓN DIM_CARRERA.

Columna	Tipo de dato	Descripción
SK_CARRERA	INTEGER	Clave subrogada de la dimensión. Esta clave se generará en los procesos ETL después de aplicar la técnica de control de cambios SCD tipo 2.
PK_CARRERA	INTEGER	Clave primaria proveniente del OLTP, identificador único de una carrera.
NOMBRE	VARCHAR (100)	Nombre de la Carrera
MODALIDAD	VARCHAR (45)	Modalidad de una carrera (Presencial, Distancia)
VALIDO_DESDE	DATE	Fecha desde la cual es válido el registro
VALIDO_HASTA	DATE	Fecha hasta cual es válido el registro
VERSION	INTEGER	Bandera que permite indicar si el registro es válido o sufrió algún cambio después de aplicar la técnica SCD tipo 2. Los posibles valores serían 1 si el registro está activo o 0 si no lo está.

2. Dimensión DIM_TITULO.

La dimensión DIM_TITULO, representa cada una de los títulos profesionales que poseen los docentes en una determinada oferta académica y carrera, esta dimensión nos permitirá realizar un análisis del nivel de titularidad docente que existe en cada una de las carreras del AEIRNNR en una determinada oferta académica.



TABLA XXVIII DIMENSIÓN DIM_TITULO.

Columna	Tipo de dato	Descripción
SK_TITULO	INTEGER	Clave subrogada de la dimensión. Esta clave se generará en los procesos ETL después de aplicar la técnica de control de cambios SCD tipo 2.
PK_TITULO	INTEGER	Clave primaria proveniente del OLTP, identificador único de la tabla título.
VALIDO_DESDE	DATE	Fecha desde la cual es válido el registro
VALIDO_HASTA	DATE	Fecha hasta cual es válido el registro
VERSION	INTEGER	Bandera que permite indicar si el registro es válido o sufrió algún cambio después de aplicar la técnica SCD tipo 2. Los posibles valores serían 1 si el registro está activo o 0 si no lo está.

3. Dimensión DIM_CATEGORIA.

La dimensión DIM_CATEGORIA, representa cada una de las categorías laborales que posee un docente, esta dimensión permitirá realizar un análisis de la población docente del AEIRNNR por categoría laboral (titular - contratado).

TABLA XXIX DIMENSIÓN DIM_CATEGORIA.

Columna	Tipo de dato	Descripción
SK_CATEGORIA	INTEGER	Clave subrogada de la dimensión. Esta clave se generará en los procesos ETL después de aplicar la técnica de control de cambios SCD tipo 2.
PK_CATEGORIA	INTEGER	Clave primaria proveniente del OLTP, identificador único de la tabla categoría.
VALIDO_DESDE	DATE	Fecha desde la cual es válido el registro
VALIDO_HASTA	DATE	Fecha hasta cual es válido el registro
VERSION	INTEGER	Bandera que permite indicar si el registro es válido o sufrió algún cambio después de aplicar la técnica SCD tipo 2. Los posibles valores serían 1 si el registro está activo o 0 si no lo está.



4. Dimensión DIM_GENERO.

La dimensión DIM_GENERO, representa el género de los estudiantes del AEIRNNR, esta dimensión permitirá realizar un análisis de la población estudiantil por género (masculino-femenino).

TABLA XXX DIMENSIÓN DIM_GENERO.

Columna	Tipo de dato	Descripción
SK_GENERO	INTEGER	Clave subrogada de la dimensión. Esta clave se generará en los procesos ETL después de aplicar la técnica de control de cambios SCD tipo 2.
PK_GENERO	INTEGER	Clave primaria proveniente del OLTP, identificador único de la tabla género.
VALIDO_DESDE	DATE	Fecha desde la cual es válido el registro
VALIDO_HASTA	DATE	Fecha hasta cual es válido el registro
VERSION	INTEGER	Bandera que permite indicar si el registro es válido o sufrió algún cambio después de aplicar la técnica SCD tipo 2. Los posibles valores serían 1 si el registro está activo o 0 si no lo está.

5. Dimensión DIM_FECHA.

La dimensión DIM_FECHA, es una de las dimensiones que siempre está presente en cada Data Mart. Esta dimensión permitirá realizar el análisis histórico del rendimiento académico dentro del AEIRNNR.

TABLA XXXI DIMENSIÓN DIM_FECHA

Columna	Tipo de dato	Descripción
SK_FECHA	INTEGER	Clave subrogada de la dimensión
OFERTA_ACADEMICA	VARCHAR(250)	Oferta académica o semestre.
PERIODO	VARCHAR (100)	Periodo lectivo o año calendario
FECHA_INICIO	DATE	Fecha inicio de clases en una Oferta Académica
FECHA_FIN	DATE	Fecha fin de clases en una Oferta Académica.



6. Dimensión DIM_ESTUDIANTE.

La dimensión DIM_ESTUDIANTE, representa cada una de los estudiantes pertenecientes al AEIRNNR a partir del año 2008 hasta la fecha, esta dimensión nos permitirá calcular la población estudiantil por oferta académica y carrera.

TABLA XXXII DIMENSIÓN DIM_ESTUDIANTE.

Columna	Tipo de dato	Descripción
SK_ESTUDIANTE	INTEGER	Clave subrogada de la dimensión. Esta clave se generará en los procesos ETL después de aplicar la técnica de control de cambios SCD tipo 2.
PK_ESTUDIANTE	VARCHAR (12)	Clave primaria proveniente del OLTP, identificador único de un estudiante (cedula).
GENERO	VARCHAR (45)	Genero del alumno (masculino - femenino)
VALIDO_DESDE	DATE	Fecha desde la cual es válido el registro
VALIDO_HASTA	DATE	Fecha hasta cual es válido el registro
VERSION	INTEGER	Bandera que permite indicar si el registro es válido o sufrió algún cambio después de aplicar la técnica SCD tipo 2. Los posibles valores serian 1 si el registro está activo o 0 si no lo está.

7. Dimensión DIM_DOCENTE

La dimensión DIM_DOCENTE, representa cada una de los docentes pertenecientes al AEIRNNR a partir del año 2008 hasta la fecha, esta dimensión permitirá conocer el total de docentes de los que dispone el AEIRNNR en la oferta actual y compararlos con ofertas anteriores.



TABLA XXXIII DIMENSIÓN DIM_DOCENTE.

Columna	Tipo de dato	Descripción
SK_DOCENTE	INTEGER	Clave subrogada de la dimensión. Esta clave se generará en los procesos ETL después de aplicar la técnica de control de cambios SCD tipo 2.
PK_DOCENTE	VARCHAR (12)	Clave primaria proveniente del OLTP, identificador único de un docente (cedula).
NOMBRES	VARCHAR (45)	Nombre del docente.
APELLIDOS	VARCHAR (45)	Apellidos del docente
VALIDO_DESDE	DATE	Fecha desde la cual es válido el registro
VALIDO_HASTA	DATE	Fecha hasta cual es válido el registro
VERSION	INTEGER	Bandera que permite indicar si el registro es válido o sufrió algún cambio después de aplicar la técnica SCD tipo 2. Los posibles valores serían 1 si el registro está activo o 0 si no lo está.

8. Tabla de hechos FACT_ANALISIS_RENDIMIENTO

Representa la información resumizada e histórica relacionada con la parte de estudiantes del AEIRNNR, dentro de esta tabla se realiza la juntura con las diferentes dimensiones pertenecientes al Data Mart de estudiantes.

TABLA XXXIV TABLA DE HECHOS FACT_ANALISIS_RENDIMIENTO.

Columna	Tipo de dato	Descripción
SK_ANALISIS_RENDIMIENTO	INTEGER	Clave primaria o identificador único de la tabla FACT_ANALISIS_RENDIMEINT O
SK_FECHA	INTEGER	Clave foránea que hace referencia a la dimensión DIM_FECHA.
SK_ESTUDIANTE	INTEGER	Clave foránea que hace



		referencia a la dimensión DIM_ESTUDIANTE.
SK_CARRERA	INTEGER	Clave foránea que hace referencia a la dimensión DIM_CARRERA.
SK_GENERO	INTEGER	Clave foránea que hace referencia a la dimensión DIM_GENERO
CALIFICACION	NUMERIC (2,2)	Calificación del estudiante en una materia
APROBADO	INTEGER	Estado: aprobado - reprobado

9. Tabla de hechos FACT_ANALISIS_DOCENTE

Representa la información sumariada e histórica relacionada con la parte de DOCENTES del AEIRNNR, dentro de esta tabla se realiza la juntura con las diferentes dimensiones pertenecientes al Data Mart de docentes.

TABLA XXXV TABLA DE HECHOS FACT_ANALISIS_DOCENTE.

Columna	Tipo de dato	Descripción
SK_ANALISIS_DOCENTE	INTEGER	Clave primaria o identificador único de la tabla FACT_ANALISIS_RENDIMIENTO
SK_FECHA	INTEGER	Clave foránea que hace referencia a la dimensión DIM_FECHA.
SK_DOCENTE	INTEGER	Clave foránea que hace referencia a la dimensión DIM_DOCENTE.
SK_CARRERA	INTEGER	Clave foránea que hace referencia a la dimensión DIM_CARRERA.
SK_TITULO	INTEGER	Clave foránea que hace referencia a la dimensión DIM_TITULO
SK_CATEGORIA	INTEGER	Clave foránea que hace referencia a la dimensión DIM_CATEGORIA



Anexo 4: Conexión base de datos con Pentaho Data Integration (PDI).

Desde la interfaz de Pentaho Data Integration se creó la conexión a través de la opción Fichero ► Nuevo ► Database Connection y a continuación se realizó las configuraciones mostradas en la figura.

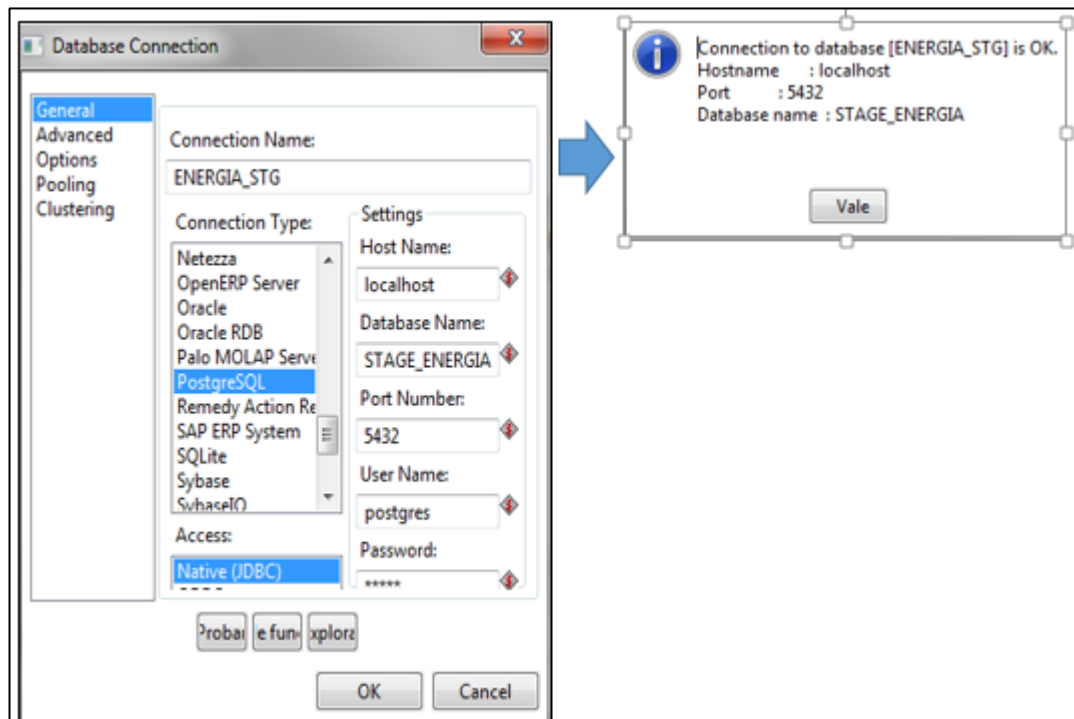


Figura 43 Conexión a BDD con PDI.



Anexo 5: ETL CARGA_OFERTAS_STAGE.

El flujo CARGA_OFERTAS_STAGE, contiene los flujos para transportar la información de cada oferta académica desde el SGAWebservices al STAGE_ENERGIA.

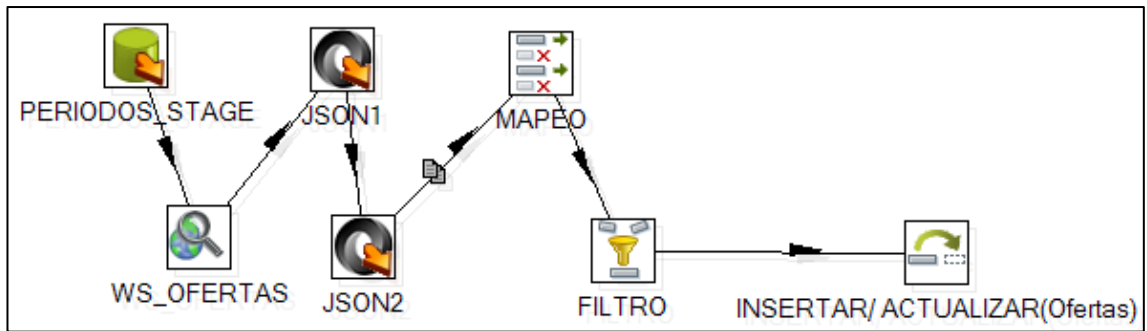


Figura 44 Transformación CARGA_OFERTAS_STAGE.

En el flujo de carga se realizó una conexión al STAGE_ENERGIA, para obtener el identificador de la tabla periodo_lectivo, que ya ha sido poblada de datos con flujos anteriores, todo esto se lo realizó con el paso PERIODOS_STAGE mediante conexión a la base de datos STAGE_ENERGIA y la siguiente consulta SQL:

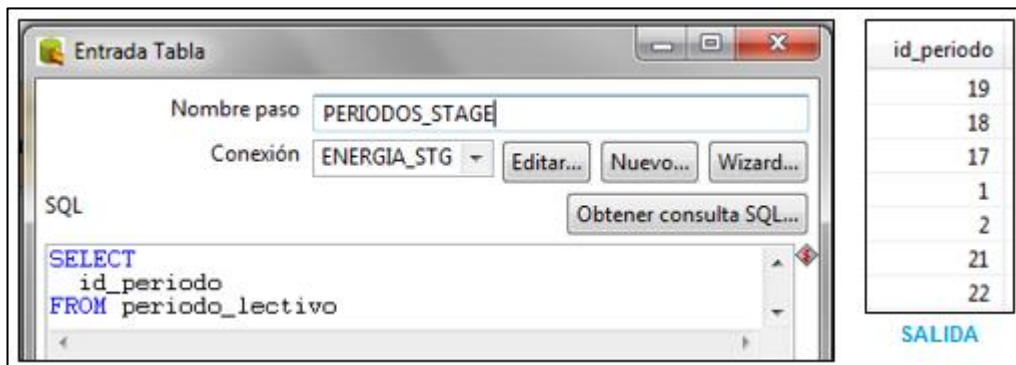


Figura 45 Configuración del paso PERIODOS_STAGE.

A continuación, mediante el paso WS_OFERTAS, se realizó una conexión al SGAWebservices, donde se invoca al método sgaws_ofertas_academicas que recibe como parámetro el id de cada periodo lectivo del paso PERIODOS_STAGE, y devuelve la información de cada oferta académica, la configuración realizada al paso WS_OFERTAS se muestra a continuación:

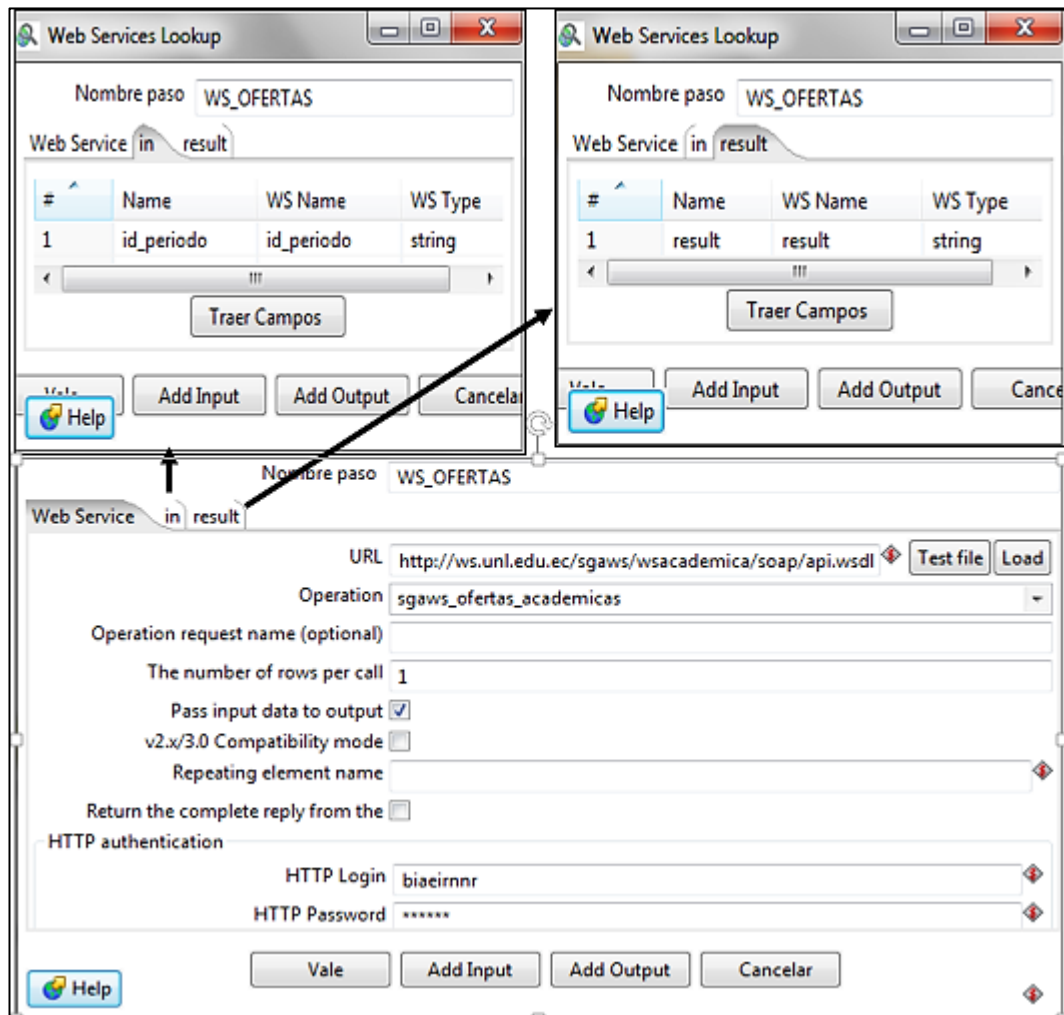


Figura 46 Configuración del paso WS_OFERTAS.

Como se puede apreciar en la Fig. 46, en la configuración se ingresa la URL: <http://ws.unl.edu.ec/sgaws/wsacademica/soap/api.wsdl>, el nombre de usuario y contraseña para conectarse al SGASWebServices y leer el WSDL (Web Services Description Language), para determinar qué funciones están disponibles en el servidor (**Load**), en este caso se ha seleccionado el método `sgaws_ofertas_academicas`, mediante la lista desplegable de la etiqueta **Operation**.

A continuación se ingresa los parámetros de entrada que recibe el método, mediante la pestaña **in**, donde se configuró el valor devuelto por el paso anterior `id_paralelo`, y mediante la pestaña **result**, se configuró el nombre de la salida que nos proporcionará el método. A continuación se muestra la salida del paso WS_OFERTAS, después de realizadas las configuraciones:



```
<result>
[[35, "Admisiones Pregrado 2010", "2010-08-09", "2010-08-18"],
[36, "Pregrado Septiembre 2010 Febrero 2011", "2010-09-11", "2011-02-23"],
[37, "Admisiones MED 2010", "2010-08-28", "2010-08-31"],
[40, "Admisiones MED 2011 (marzo)", "2011-02-12", "2011-02-13"],
[42, "Pregrado Marzo 2011 - Julio 2011 ", "2011-03-09", "2011-08-05"],
[39, "Cursos Especiales", "2010-09-01", "2012-01-01"],
[41, "Internado 2010-2011", "2010-09-01", "2011-08-31"]]
</result>
```

Figura 47 Salida del paso WS_OFERTAS.

La salida del paso WS_OFERTAS mostrada en la figura anterior, se encuentra en formato JSON, y está compuesta de un arreglo que contiene como elementos un conjunto de arreglos, todos ellos separados por una coma. Para extraer los arreglos de manera individual se utilizó el paso JSON1, en donde mediante la lista desplegable de la etiqueta Get Source From Field, se toma nuevamente la variable result y se configura el Path (\$.*) en la pestaña Field, para extraer cada arreglo de manera individual y almacenarlo en la variable ofertas.

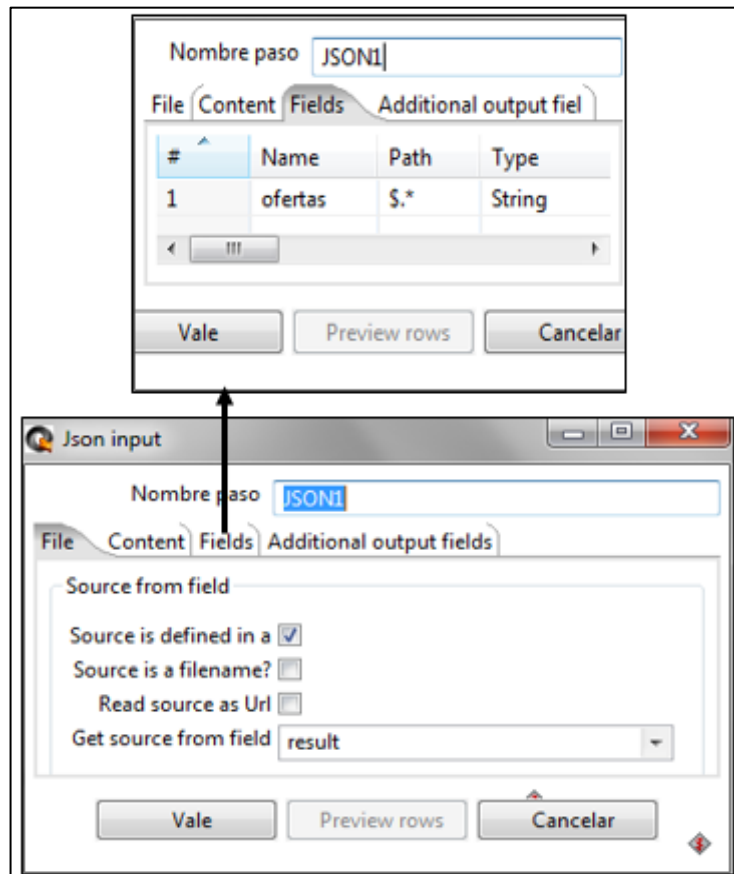


Figura 48 Configuración del paso JSON1.



ofertas
[35,"Admisiones Pregrado 2010","2010-08-09","2010-08-18"]
[36,"Pregrado Septiembre 2010 Febrero 2011","2010-09-11","2011-02-23"]
[37,"Admisiones MED 2010","2010-08-28","2010-08-31"]
[40,"Admisiones MED 2011 (marzo)","2011-02-12","2011-02-13"]
[42,"Pregrado Marzo 2011 - Julio 2011 ","2011-03-09","2011-08-05"]
[39,"Cursos Especiales","2010-09-01","2012-01-01"]
[41,"Internado 2010-2011","2010-09-01","2011-08-31"]

Figura 49 Salida del paso JSON1.

En la Fig. 49, el salto JSON1 nos retorna como salida los arreglos de manera individual, ahora se extraerá los datos de manera individual de cada arreglo, en este caso el id, descripción, fecha de inicio y fecha de fin de cada oferta académica, mediante el paso JSON2, cuya configuración se muestra a continuación:

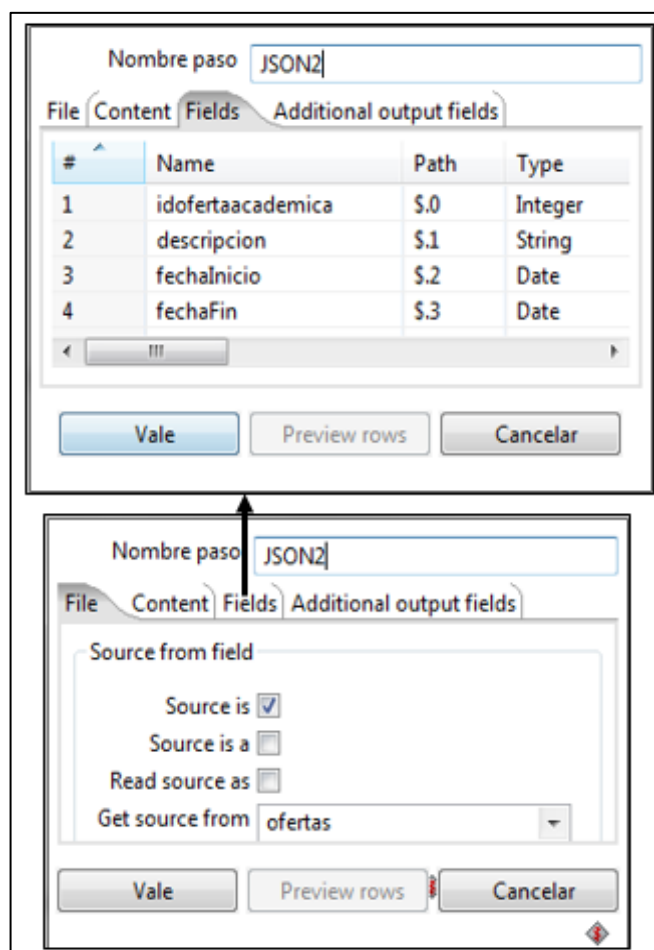


Figura 50 Configuración del paso JSON2.



idofertaacademica	descripcion	fechaInicio	fechaFin
35	Admisiones Pregrado 2010	2010-08-09	2010-08-18
36	Pregrado Septiembre 2010 Febrero 2011	2010-09-11	2011-02-23
37	Admisiones MED 2010	2010-08-28	2010-08-31
40	Admisiones MED 2011 (marzo)	2011-02-12	2011-02-13
42	Pregrado Marzo 2011 - Julio 2011	2011-03-09	2011-08-05
39	Cursos Especiales	2010-09-01	2012-01-01
41	Internado 2010-2011	2010-09-01	2011-08-31

Figura 51 Salida del paso JSON2.

Finalmente esta información es llevada a la tabla oferta_academica de la base de datos STAGE_ENERGIA, mediante el paso INSERTAR/ACTUALIZAR (Ofertas), que nos permitió realizar cargas incrementales mediante la clave primaria de la misma tabla.

Insertar/Actualizar

Nombre de paso: INSERTAR/ ACTUALIZAR(Ofertas)

Conexión: ENERGIA_STG [Editar... Nuevo... Wizard...]

Esquema destino: [Examinar...]

Tabla destino: oferta_academica [Examinar...]

Tamaño transacción (commit): 100

No realizar actualizaciones:

La clave(s) para realizar búsqueda de valor(es):

#	Campo de tabla	Comparador	Campo1	Cam2
1	id_oferta	=	idofertaacademica	

[Obtener Campos]

Campos de actualización:

#	Campo de tabla	Campo de Flujo	Actualizar
1	id_oferta	idofertaacademica	Y
2	descripcion	descripcion	Y
3	fecha_inicio	fecha_inicio	Y
4	fecha_fin	fecha_fin	Y
5	id_periodo	id_periodo	Y

[Obtener campos de actualización] [Edit mapping]

[Help] [Vale] [Cancelar] [SQL]

Figura 52 Configuración del paso INSERTAR/ ACTUALIZAR (Ofertas).



Anexo 6: Configuración de transformaciones del trabajo UNIR_TRANSF_ESTUDIANTE.

1. ETL CARGA_DIM_CARRERA_ESTUDIANTE.

El flujo DIM_CARRERA_ESTUDIANTE, contiene los flujos de datos para transportar la información de cada carrera del STAGE_ENERGIA al Data Mart DM_ESTUDIANTES_AEIRNNR.



Figura 53 Transformación CARGA_DIM_CARRERA_ESTUDIANTE.

En el flujo de carga se realizó una conexión al STAGE_ENERGIA, mediante el paso DATOS_CARRERA, y se extrae los campos (id_carrera, nombre) de la tabla carreras, para luego ser convertidos en mayúsculas a través del paso CONVERTOR_MAY, y finalmente llevarlos a la tabla dim_carrera del Data Mart DM_ESTUDIANTES_AEIRNNR a través del paso DIM_CARRERA.



Figura 54 Configuración del paso DATOS_CARRERA.



Búsqueda/Actualización en Dimensión

Nombre paso DIM_CARRERA

¿Actualizar la dimensión? Permite actualizar dimensión

Conexión DM_AREA Conexión al DM_AEIRNNR [Editar...] [Nuevo...] [Wizard...]

Esquema destino [Examinar...]

Tabla destino dim_carrera dimensión carrera [Examinar...]

Tamaño transacción 100

Enable the cache?

Pre-load the cache?

Cache size in rows (0 = cache all) 5000

Claves Campos

Campos clave (para buscar fila en dimensión):

#	Campo de Dimensión	Campo en flujo
1	pk_carrera	pk_carrera Clave primaria de la tabla carrera del STAGE_AEIRNNR

Campo de clave técnica sk_carrera Clave subrogada [Nuevo nombre]

Creación de clave técnica

Utilizar máximo tabla + 1

Utilizar secuencia

Utilizar campo auto-incrementativo

Campo de Versión version versión de la entrada dim_carrera (0=no valido, 1=valido)

Campo Fecha flujo

Campo inicio rango fecha valido_desde Año mín. 2008

Use an alternative start date? <Select Option>

Campo final rango fecha valido_hasta Año máx. 2020

[Vale] [Cancelar] [Obtener Campos] [SQL]

[Help]

Figura 55 Configuración del paso DIM_CARRERA.

Dentro del paso DIM_CARRERA, se aplicó la técnica SCD tipo 2 y el concepto de claves subrogadas, como se puede apreciar en la Fig. 55 se establece una conexión al DM_ESTUDIANTES_AEIRNNR, para luego seleccionar la tabla dim_carrera (tabla destino), posteriormente se toma el campo pk_carrera que es la llave primaria de la tabla carrera proveniente del paso anterior, sk_carrera hace referencia a la clave subrogada de tipo numérico que el paso genera automáticamente. El campo valido_desde se refiere a la fecha desde la cual es válido el registro, mientras que el campo valido_hasta se refiere a la fecha hasta la cual es válido el registro, todo esto es controlado mediante el campo versión que nos permitirá indicar si el registro es

válido o sufrió algún cambio después de aplicar la técnica SCD tipo 2. Los posibles valores serían 1 si el registro está activo, o 0 si no lo está.

2. ETL DIM_FECHA_ESTUDIANTE.

El flujo DIM_FECHA_ESTUDIANTE, contiene los procesos ETL que permiten poblar la dimensión dim_fecha, como se mencionó anteriormente el nivel de jerarquía de esta dimensión está definida en semestres (ofertas académicas) y años (periodos lectivos).

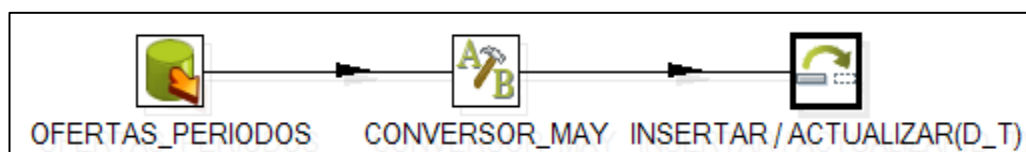


Figura 56 Transformación DIM_FECHA_ESTUDIANTE.

En el flujo de carga se realizó la conexión al STAGE_ENERGIA, mediante el paso OFERTAS_PERIODOS, para luego mediante la juntura de las tablas relacionadas oferta_academica y periodo_lectivo, tomar cada uno de los atributos necesarios para llenar la tabla dim_fecha, la clave subrogada para dicha dimensión se la construyó a partir del atributo fecha_inicio, de donde se extrae el año y el mes de la misma, permitiéndonos así crear un atributo único para dicha tabla, todo esto se lo realizó mediante el código SQL mostrado en la Fig. 57.

```
select to_number(to_char(fecha_inicio,'yyyyMM'),'999999') as sk_fecha,  
o.id_oferta as pk_oferta_academica, o.descripcion as oferta_academica,  
p.descripcion as periodo, o.fecha_inicio as fecha_inicio, o.fecha_fin  
as fecha_fin from oferta_academica o inner join periodo_lectivo p  
on o.id_periodo=p.id_periodo
```

Figura 57 Configuración del paso OFERTAS_PERIODOS.

Una vez implementada la consulta SQL en el paso OFERTAS_PERIODOS, se convirtió en mayúsculas el valor almacenado en el campo oferta_academica mediante el paso CONVERSION_MAY, para luego ser llevada a la tabla dim_fecha a través del paso INSERTAR / ACTUALIZAR (D_T).

3. ETL HECHOS_ESTUDIANTE.

El siguiente proceso después de cargar las diferentes dimensiones, consiste en poblar la tabla de hechos, para ello se configuró el flujo HECHOS_ESTUDIANTE, que es el que contiene los procesos ETL para dicha actividad.

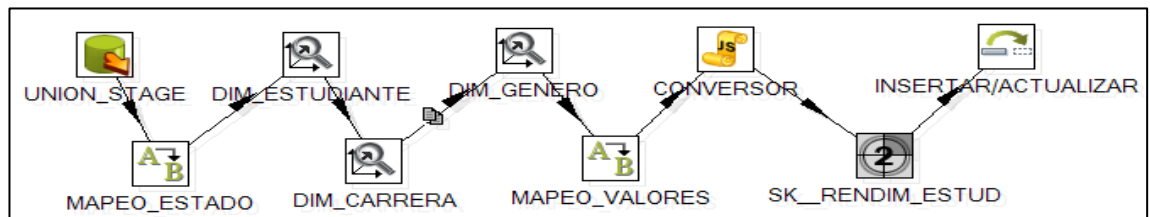


Figura 58 Transformación HECHOS_ESTUDIANTE.

En el flujo HECHOS_ESTUDIANTE, mediante el paso UNION_STAGE se realizó la conexión a la tabla calificaciones del STAGE_ENERGIA, que es donde se almacena las notas de los estudiantes y otros datos necesarios para poblar la tabla de hechos. Una vez establecida la conexión, mediante código SQL se realizó la juntura de la tabla calificaciones y las tablas relacionadas con la misma, para luego extraer las claves primarias de cada tabla, la calificación y el estado del estudiante (aprobado - reprobado). La sentencia SQL empleada para dicho propósito se muestra a continuación:

```
select to_number(to_char(fecha_inicio, 'yyyyMM'), '999999') as sk_fecha, car.id_carrera  
as pk_carrera, e.ci as pk_estudiante, de.id_genero as pk_genero, c.nota as calificacion,  
c.estado from oferta_academica o inner join calificaciones c on o.id_oferta=c.id_oferta  
inner join carrera car on car.id_carrera=c.id_carrera inner join estudiante e on e.ci=c.ci  
inner join datos_estudiante de on de.ci=e.ci
```

Figura 59 Configuración del paso UNION_STAGE.

De los resultados devueltos por la consulta SQL, se realizó el mapeo de valores de la columna aprobado mediante el paso MAPEO_APREP, es decir en caso de contener el valor EstadoMatriculaReprobada se asigna el valor de 0, caso contrario se asigna el valor de 1, luego a este mismo valor se lo convirtió a entero mediante el paso CONVERSION, en donde mediante la utilización de código JavaScript se realizó la conversión antes mencionada. A continuación se muestra las configuraciones realizada en dichos pasos.

```
Script1 ✖  
//Script here  
  
var aprobado2=parseInt(aprobado.getString())
```

Figura 60 Código JavaScript empleado en el salto CONVERSION.



Nombre de paso : MAPEO_APREP

Nombre de campo : aprobado

Nombre de campo :

Default upon :

Valores de campo:

#	Valor origen	Valor destino
1	EstadoMatriculaReprobada	0
2	EstadoMatriculaAprobada	1

Buttons: Help, Vale, Cancelar

Figura 61 Configuración del paso MAPEO_APREP.

Los pasos DIM_ESTUDIANTE, DIM_CARRERA, DIM_GENERO, permiten extraer los datos de cada dimensión, por ejemplo para obtener los campos de la dimensión carrera, lo que se realizó en el salto DIM_CARRERA fue establecer una conexión al DM_ESTUDIANTES_AEIRNNR, para luego seleccionar la tabla dim_carrera, a continuación el paso realiza internamente una búsqueda dentro de la tabla dim_carrera mediante el campo pk_carrera proveniente del paso anterior, una vez que lo encuentra extrae la clave subrogada de la dimensión carrera y la retorna como salida para poder ser almacenada en la tabla fact_rend_estudiante y así referenciar a la dimensión carrera.

La configuración realizada en el paso DIM_CARRERA, es la que siguen los pasos posteriores, variando solo el nombre de los campos y la dimensión, razón por la cual no se entrará en detalle en los demás pasos, a excepción del paso INSERTAR/ACTUALIZAR_HECHOS cuya estructura es diferente.



Búsqueda/Actualización en Dimensión

Nombre paso: DIM_CARRERA

¿Actualizar la dimensión?

Conexión: DM_AREA

Esquema destino:

Tabla destino: dim_carrera

Tamaño transacción: 100

Enable the cache?

Pre-load the cache?

Cache size in rows (0 = cache all): 5000

Claves Campos

Campos clave (para buscar fila en dimensión):

#	Campo de Dimensión	Campo en flujo
1	pk_carrera	pk_carrera

Campo de clave técnica: sk_carrera

Creación de clave técnica

- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

Campo de Versión: version

Campo Fecha flujo:

Campo inicio rango fecha: valido_desde

Use an alternative start date? <Select Option>

Campo final rango fecha: valido_hasta

Vale Cancelar Obtener Campos

Help

Figura 62 Configuración del paso DIM_CARRERA.

Después de realizar las configuraciones de los pasos de cada una de las dimensiones, se tomó los datos devueltos por cada uno de ellos y se los llevó finalmente a la tabla fact_rend_estudiante mediante el paso INSERTAR/ACTUALIZAR_HECHOS, cuya configuración se muestra a continuación.



Nombre de paso: INSERTAR/ACTUALIZAR_HECHOS

Conexión: DM_ESTUDIANTE [Editar...] [Nuevo...] [Wizard...]

Esquema destino: [Examinar...]

Tabla destino: fact_rend_estudiante [Examinar...]

Tamaño transacción (commit): 100

No realizar actualizaciones:

La clave(s) para realizar búsqueda de valor(es):

#	Campo de tabla	Comparador	Campo1	Obtener Campos
1	sk_rend_estudiante	=	sk_rend_estudiante	

Campos de actualización:

#	Campo de tabla	Campo de Flujo	A	Obtener campos de actualización
1	sk_rend_estudiante	sk_rend_estudiante	Y	
2	sk_fecha	sk_fecha	Y	
3	sk_estudiante	sk_estudiante	Y	
4	sk_carrera	sk_carrera	Y	
5	sk_genero	sk_genero	Y	
6	calificacion	calificacion	Y	
7	estado	aprobado2	N	

[Help] [Vale] [Cancelar] [SQL] [Edit mapping]

Figura 63 Configuración del paso INSERTAR/ACTUALIZAR_HECHOS.



Anexo 7: Reportes con Pentaho Report Designer.

Pentaho Report Designer está disponible en la dirección: <http://sourceforge.net/projects/pentaho/files/Report%20Designer/>. Luego de descargar el archivo deberá descomprimirse y copiar la carpeta report-designer en el directorio C:\pentaho, que ha sido creado para copiar cada una de los módulos de Pentaho.

Una vez que se ha copiado la carpeta report-designer en la ruta de instalación de Pentaho, se deberá ejecutar el archivo report-designer.bat, el mismo que esta contenido dentro de la misma carpeta.

Para iniciar un reporte, se lo puede hacer usando el Report Wizard y mediante la opción New, con la diferencia de que el primero nos proporciona plantillas personalizadas.

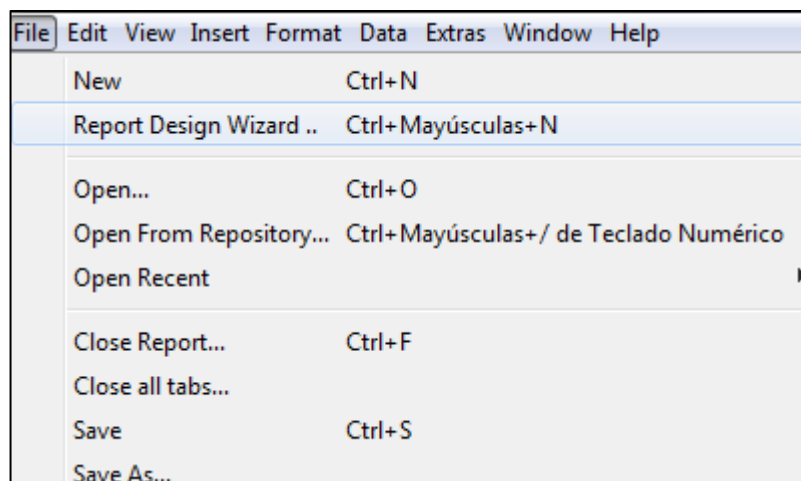


Figura 64 Generar reporte con Report Wizard.

Una vez seleccionada la opción Report Wizard, se nos mostrará las siguientes opciones:

- **Look and Feel:** en esta área podemos definir como se verá el reporte.
- **Template:** permite seleccionar plantillas determinadas.
- **Report Document:** marcándolo podremos cargar un archivo con la especificación de configuración para el reporte.

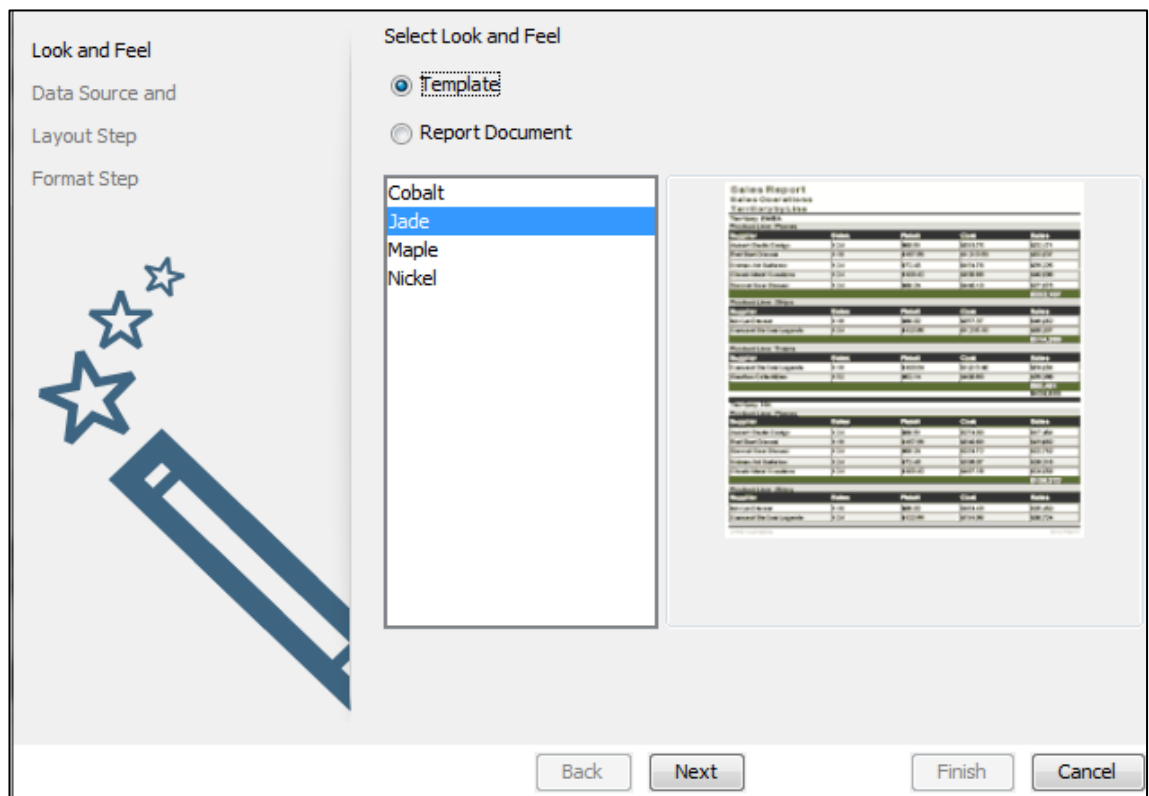


Figura 65 Plantillas Report Wizard.

Presionamos **Next** y en la siguiente interfaz seleccionamos Sample Query, a continuación se activa en la parte superior derecha el icono de un lápiz, el mismo que nos permite crear la conexión a la base de datos.

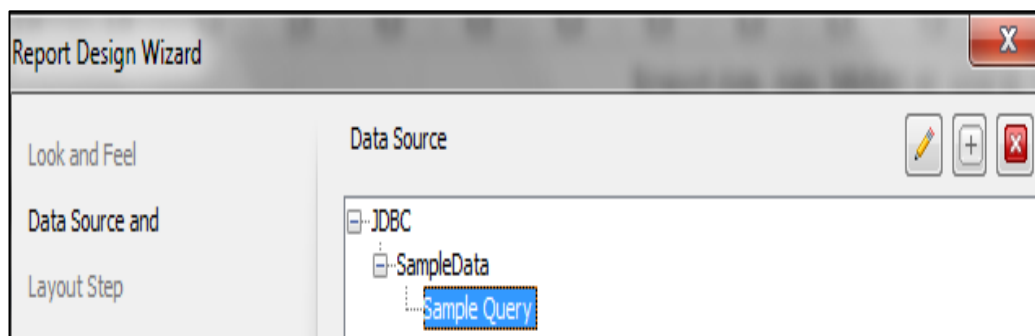


Figura 66 Conexión a base de datos con Report Wizard.

En la siguiente interfaz, se debe seleccionar el icono de lápiz que está en la parte superior izquierda, en donde se nos permitirá realizar las configuraciones para la conexión a la base de datos, en este caso al Data Mart de docentes.

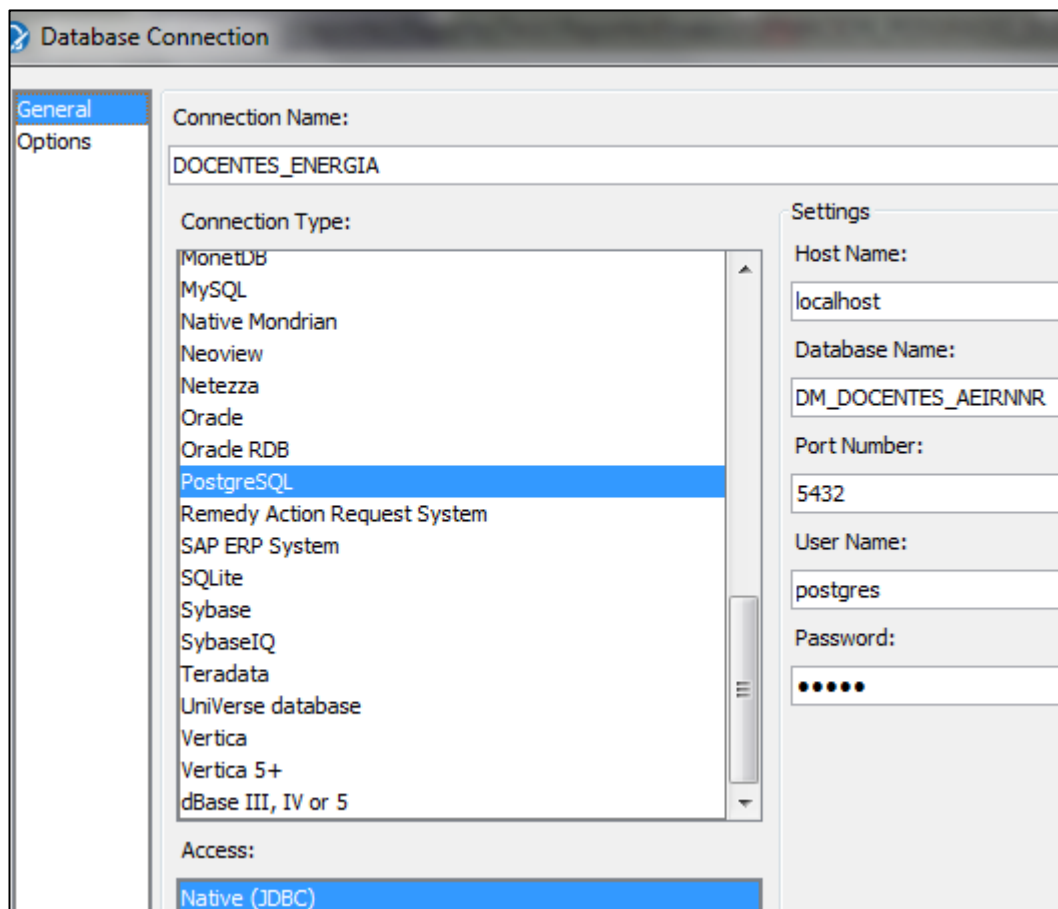


Figura 67 Configuración de acceso a base de datos con Pentaho Report Designer.

Presionamos OK, luego nos ubicamos en el nombre de nuestra conexión y presionamos en el signo más que se encuentra en la parte superior derecha, para agregar nuestro primer Query, le asignamos un nombre y en la parte superior del mismo asignamos nuestra consulta SQL. En este caso, se ha diseñado una consulta que nos permita extraer los docentes con título profesional de cuarto nivel, en base a los parámetros oferta académica y carrera. Las consultas para obtener cada una de las ofertas y carreras, que serán los parámetros del reporte, se las diseñaron dentro de esta misma interfaz, a continuación se detallan las consultas SQL creadas.

```
1 SELECT
2     "public"."dim_fecha"."oferta_academica"
3 FROM
4     "public"."dim_fecha"
```

Figura 68 Consulta SQL ofertas.



```
1 SELECT df.oferta_academica as oferta, car.nombre as carrera,  
2 count (DISTINCT fa.sk_docente) as magister, dt.pk_titulo as  
3 titulo FROM fact_analisis_docente fa inner join dim_fecha df on df.sk_fecha=fa.sk_fecha  
4 INNER JOIN dim_carrera car on fa.sk_carrera=car.sk_carrera  
5 inner join dim_docente dd on dd.sk_docente=fa.sk_docente  
6 INNER JOIN dim_titulo dt on dt.sk_titulo=fa.sk_titulo  
7 GROUP BY df.sk_fecha, car.sk_carrera, dt.pk_titulo  
8 HAVING (dt.pk_titulo like '%MAESTRIA%' OR dt.pk_titulo  
9 like '%MÁSTER%' or dt.pk_titulo like '%MAGISTER%'  
10 or dt.pk_titulo like '%MAGÍSTER%' or dt.pk_titulo  
11 like '%MG. SC.%' or dt.pk_titulo like '%MAESTRIA%' or dt.pk_titulo  
12 like '%MASTER%' or dt.pk_titulo like '%BACHELOR%'  
13 and dt.pk_titulo like '%DOCTOR%' OR dt.pk_titulo like  
14 '%DOCTORA%' OR dt.pk titulo like '%DR%')  
15 AND df.oferta_academica=${ofertas_a} AND car.nombre=${carreras a} Parametros
```

Figura 69 Consulta SQL formación_posgrado.

```
1 SELECT  
2 "public"."dim_carrera"."nombre"  
3 FROM  
4 "public"."dim_carrera"  
5 WHERE  
6 "public"."dim_carrera"."nombre" <> 'N/D'
```

Figura 70 Consulta SQL carreras.

Después de configurar las consultas SQL, se procede a personalizar el reporte de acuerdo a la platilla seleccionada anteriormente. Para crear los parámetros, se lo realizó a través de la pestaña Data, ubicada en la parte superior derecha del reporte.

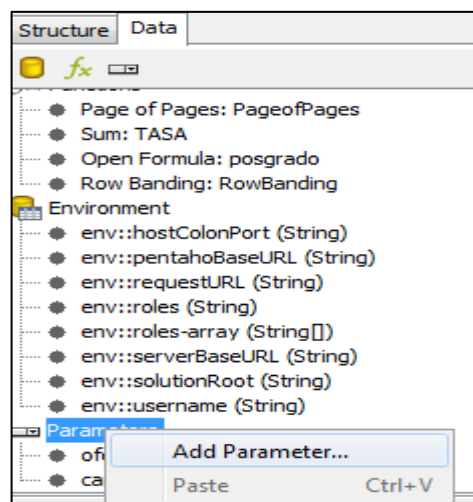


Figura 71 Creación de parámetros en Report Designer.



A continuación se le agrego nombre al parámetro, y se le asigno uno de los Querys definidos anteriormente.

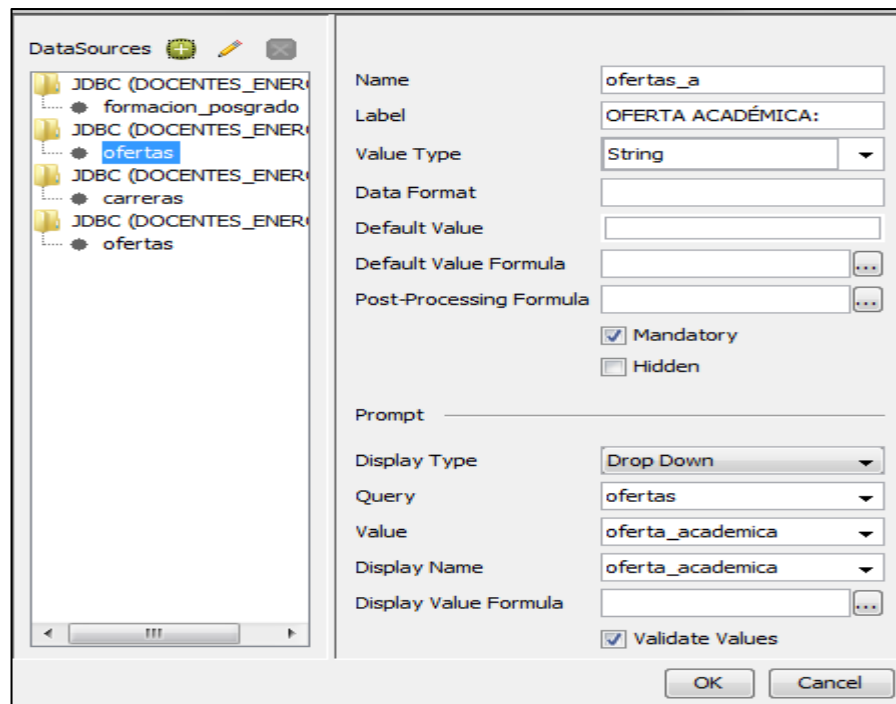


Figura 72 Configuración de parámetros con Report Designer.

Finalmente después de terminado el reporte, se lo publicó al servidor de Pentaho para que este pueda ser visualizado por el usuario. Para publicar un reporte se lo realiza desde la opción: File ► Publish, donde se nos mostrará la siguiente pantalla:

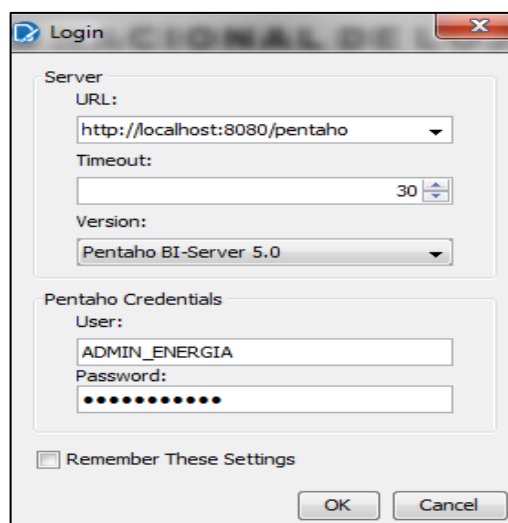


Figura 73 Publicar reporte con Report Designer.



Después de ingresar el nombre de usuario y contraseña, a continuación se elige la ruta donde se va a copiar el reporte dentro del servidor, cabe recalcar que para realizar este proceso debemos asegurarnos que el servidor de Pentaho este activo.

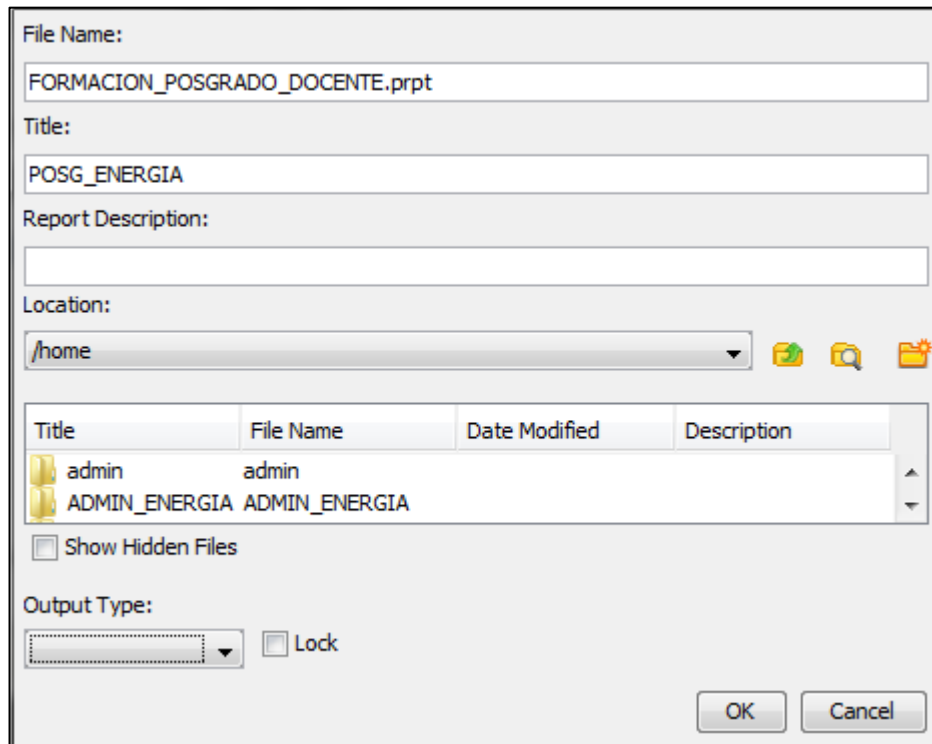


Figura 74 Agregar ruta de ubicación del archivo con PRD.

Finalmente presionamos Ok, y si todo se ha configurado correctamente se nos mostrará el siguiente mensaje, donde nos indica que el reporte se ha publicado correctamente, y con esto ya podremos visualizarlo desde el servidor de Pentaho.

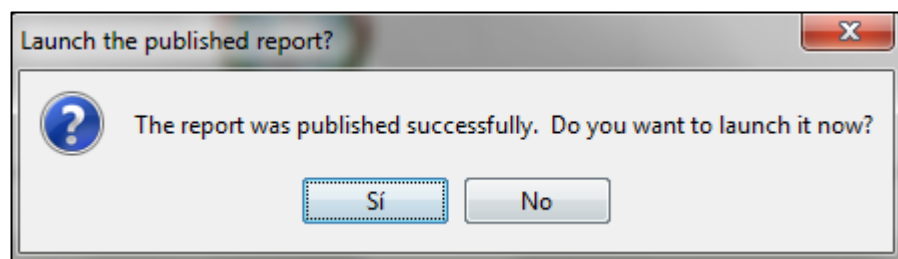


Figura 75 Confirmación de publicación de reporte con PRD.



Anexo 8: Creación de cubos OLAP con Pentaho Schema Workbench.

Pentaho Schema Workbench se encuentra disponible en: <http://sourceforge.net/projects/mondrian/files/schema%20workbench/3.6.1-stable/>. Una vez descargado se debe descomprimir y guardar la carpeta schema-workbench en la ruta de instalación de Pentaho. A continuación se debe ejecutar el archivo workbench.bat para acceder a la interfaz de la aplicación.

Una vez que ingresamos a la aplicación, se podrá establecer la conexión a la base de datos mediante: Options ► Connection, donde se configura los siguientes parámetros:

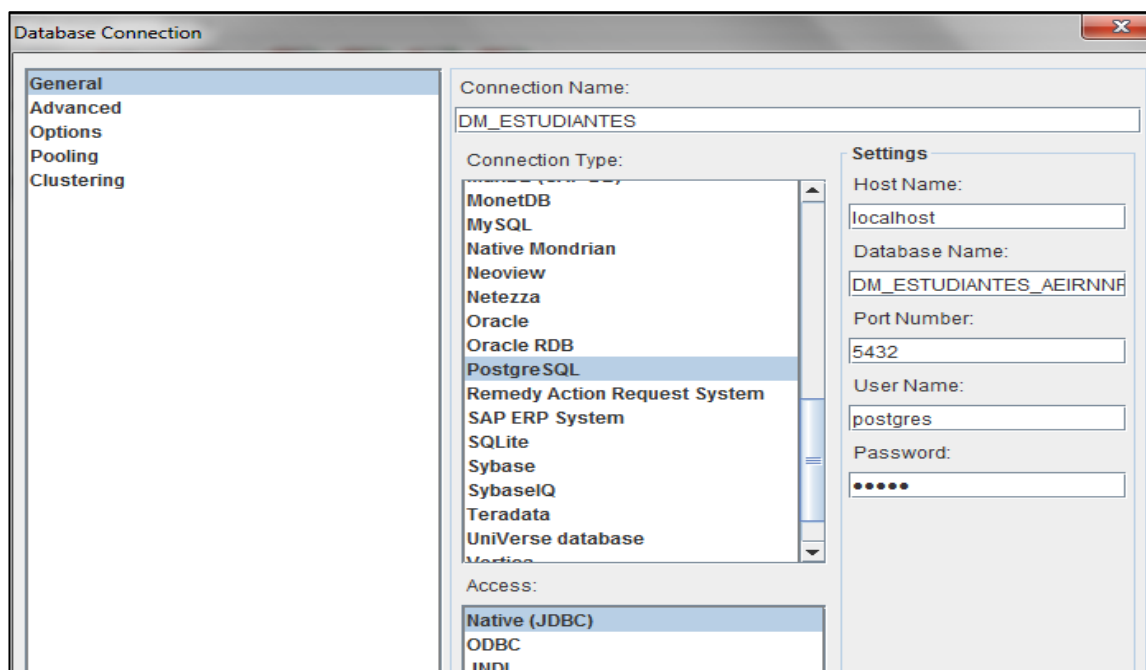


Figura 76 Acceso a base de datos con PSW.

Después de seleccionar el Data Mart de estudiantes como fuente de datos, creamos un esquema en el menú File ► New ► Schema.

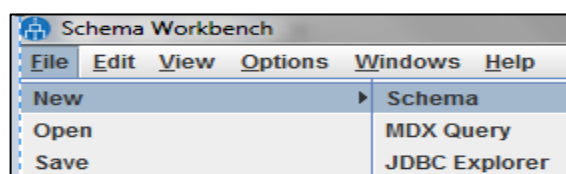


Figura 77 Creación de un esquema con PSW.



Una vez definido el esquema de nombre RENDIMIENTO_ACADEMICO, se asignó un cubo con nombre CUBO_RENDIMIENTO.

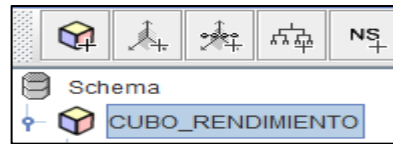


Figura 78 Creación de un cubo con PSW.

Presionando clic derecho en el cubo creado, seleccionamos la opción Add Table para agregar la tabla de hechos al cubo.

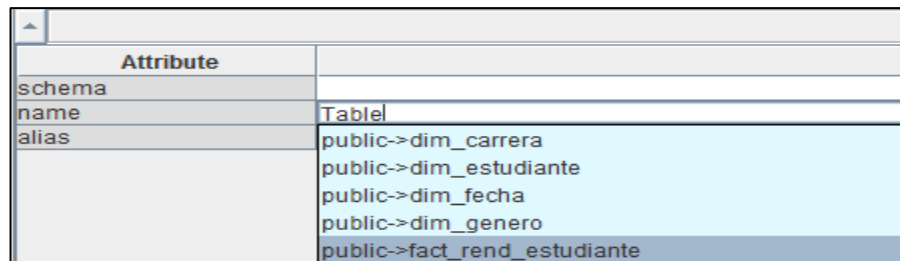


Figura 79 Agregar tabla de hechos a un cubo con PSW.

Para añadir dimensiones al cubo, presionamos clic derecho sobre el cubo y seleccionamos la opción “Add Dimension”, es aquí donde se hará referencia a las tablas de dimensiones que se quieran utilizar. Este elemento también tiene en su interior otras estructuras.

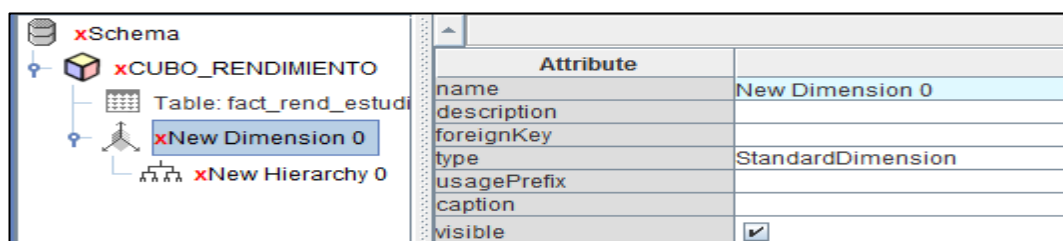


Figura 80 Crear dimensión en un cubo con PSW.

Como se puede apreciar, ha surgido un nuevo elemento “New Dimension 0”, el mismo que nos permitirá referenciar las tablas de dimensiones. Para seleccionar la tabla dimensión presionamos clic derecho sobre “New Hierarchy 0” y seleccionamos la opción “Add Table” y en el campo “name” elegimos la dimensión requerida, en este caso la dimensión dim_fecha.

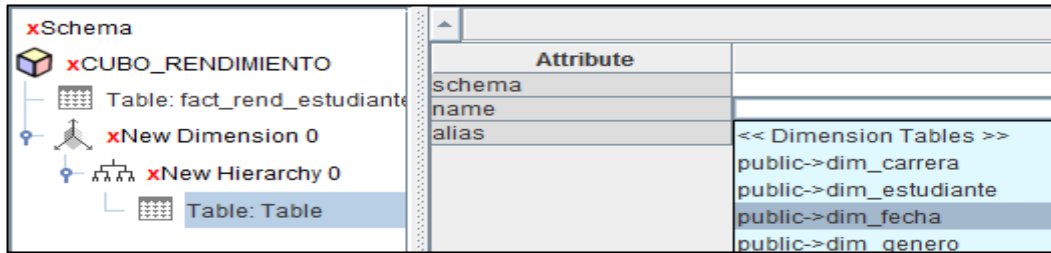


Figura 81 Agregar dimensión a un cubo con PSW.

Luego nos ubicamos en “New Hierarchy 0” y en los campos “name” y “allMemberName” ponemos FECHA, este será el título de los valores que saldrán en el reporte. En el campo primaryKey, seleccionamos de los campos que tenga esta tabla dimensión, aquel que sea la llave primaria, en este caso sk_fecha.

Attribute	
name	FECHA
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	FECHA
allMemberCaption	
allLevelName	
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	sk_fecha
caption	
visible	<input checked="" type="checkbox"/>

Figura 82 Agregar campos a mostrar en una jerarquía con PSW.

Una vez definidas las dimensiones con sus respectivas jerarquías, se agregó el cálculo de medidas al cubo, para ello nos posicionamos en el cubo, presionamos clic derecho y seleccionamos “AddMeasure”. En “name” se debe asignar el nombre del valor a calcular, en “agregator” seleccionamos la operación de sumarización que se desea realizar, en nuestro caso AVG, ya que se desea calcular el promedio de notas por oferta académica y carrera. Finalmente en “column”, seleccionamos el campo calificación, que será el hecho al que le vamos a realizar el cálculo seleccionado anteriormente.

Attribute	
name	PROMEDIOS
description	
aggregator	avg
column	calificacion
formatString	
datatype	Numeric
formatter	
caption	
visible	<input checked="" type="checkbox"/>

Figura 83 Cálculo de indicadores con PSW.



Después de asignar al cubo cada una de las dimensiones y medidas, se lo publicó al servidor Pentaho BIServer, para que este pueda ser visualizado por el usuario. Para publicar el cubo se lo realizó mediante File ► Publish, donde se deberá asignar el usuario y contraseña para su respectiva publicación.

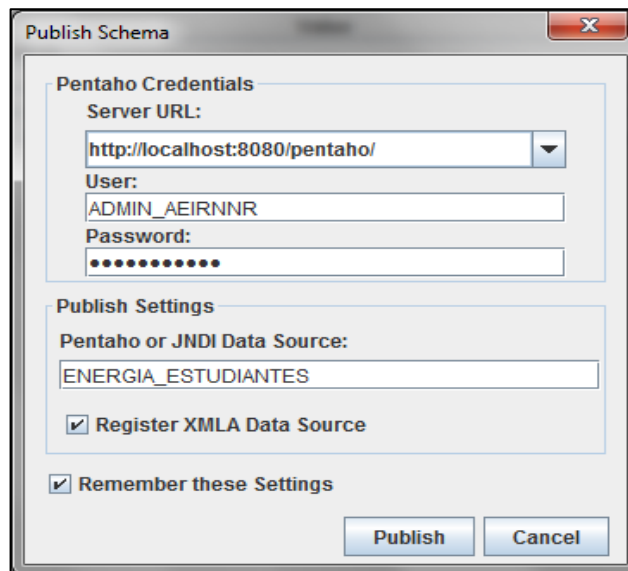


Figura 84 Publicar cubo con *Pentaho Schema Workbench*.



Anexo 9: Estructura Académica del AEIRNNR.

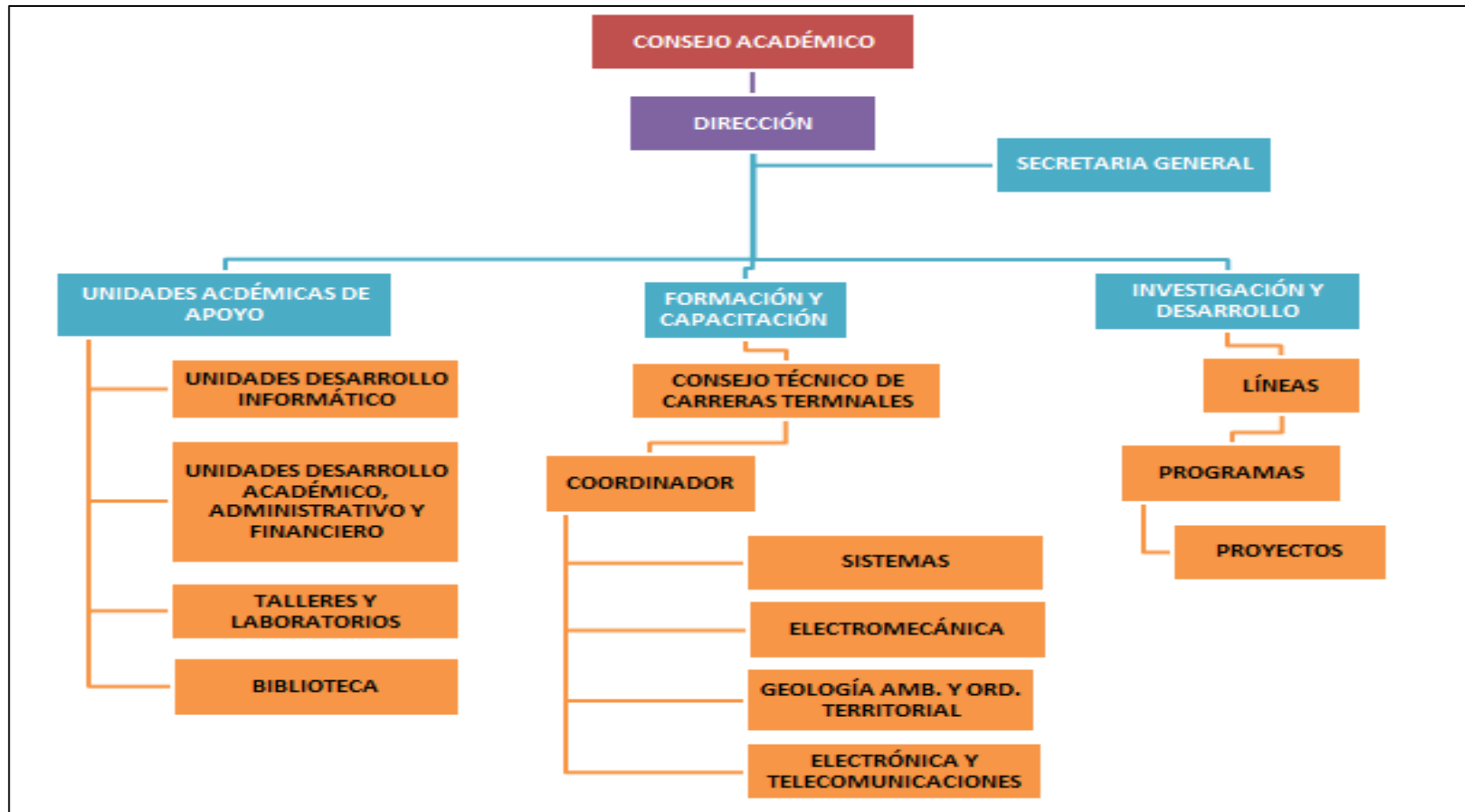


Figura 85 Estructura académica del AEIRNNR.



Descripción Estructura Académica.

1. **Consejo Académico:** es el encargado de fijar políticas y lineamientos para el funcionamiento del área, el Consejo Académico está formado por el director del área y los coordinadores de cada carrera.
2. **Director:** es el encargado de ejecutar acuerdos o resoluciones tomadas por el Consejo Académico.
3. **Secretaría General:** sirve de apoyo directo a la dirección.
4. **Unidades Académicas de Apoyo:** sirven de apoyo en la actividad práctica de la formación de estudiantes.
5. **Formación y Capacitación:** se encarga de la parte académica de cada carrera.
6. **Investigación y desarrollo:** son líneas, programas y proyectos que se podrían ejecutar dentro del área.



Anexo 10: Pentaho BIServer.

Pentaho BI Server está disponible en la dirección: <http://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/5.0.1-stable/>. Luego de descargar el archivo deberá descomprimirse y copiar la carpeta BIServer-ce en el directorio C:\pentaho.

Una vez que se ha copiado la carpeta report-designer en la ruta de instalación de Pentaho, se deberá ejecutar el archivo start-pentaho.bat contenido dentro de la misma carpeta.

A continuación, se accede a la aplicación mediante un navegador a través de la ruta: localhost: 8080, donde se nos mostrará la interfaz de logeo para acceder a la misma.



Figura 86 Consola de login de Pentaho BIServer.

Inicialmente, se puede acceder con el usuario que viene por defecto cuyos datos son:
Usuario: admin, Password: password.

Después de ingresar los campos de login que la aplicación trae por defecto, se nos muestra la consola de administración de Pentaho, cuya interfaz es la que se muestra a continuación:

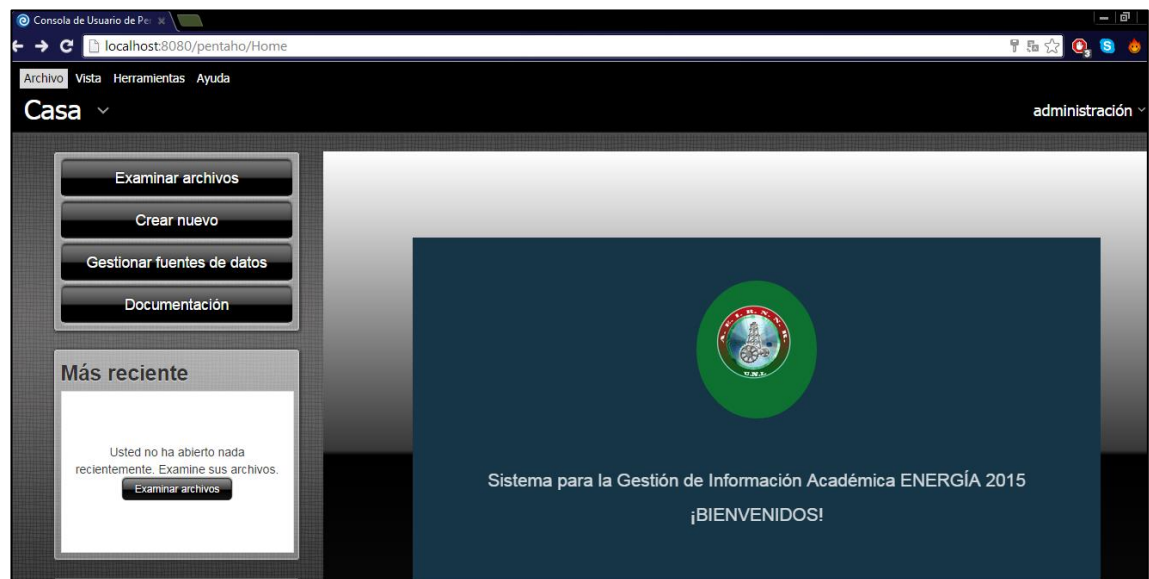


Figura 87 Consola de Administración de Pentaho BServer.

Una vez dentro de la consola de administración, se puede acceder a cada uno de los reportes y cubos OLAP, para ello se debe realizar las siguientes configuraciones:

1. Conexión a Bases de Datos.

Para la conexión a las bases de datos, se lo realizó mediante la opción Manage Data Sources, en donde se nos muestra la interfaz para realizar la conexión a la base de datos.

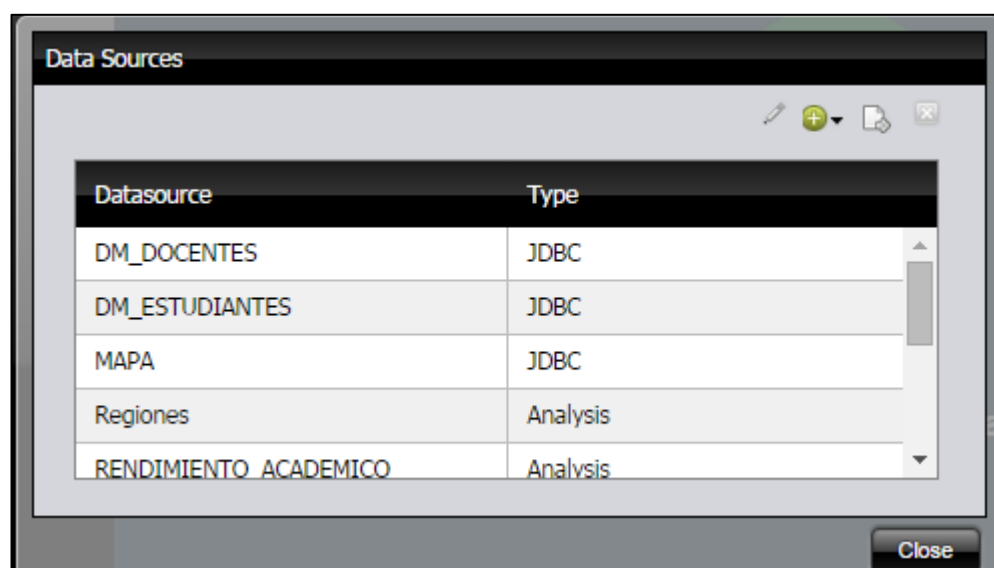


Figura 88 Conexión a BDD con Pentaho BServer.



Presionamos Add (+) ► JDBC, para seleccionar la base de datos con la que se va a trabajar, los parámetros a configurar son los siguientes:

Field	Value
Connection Name	DM_DOCENTES
Database Type	PostgreSQL
Access	Native (JDBC)
Host Name	localhost
Database Name	DM_DOCENTES_AEIRNNR
Port Number	5432
User Name	postgres
Password

Figura 89 Configuración de BDD con Pentaho BISEVER.

La configuración a la base de datos es importante, ya que al momento de importar los cubos OLAP desde la herramienta Pentaho Schema Workbench, se deberá elegir la base de datos con la que está relacionado el cubo.

2. Instalación de plugins.

Para la instalación de plugins, se lo realizó mediante la opción: HOME ► Marketplace, donde se nos mostrará una lista de plugins que la comunidad Pentaho proporciona, se debe mencionar que la disponibilidad de internet es importante al momento de realizar la instalación de plugins mediante el Marketplace de Pentaho.



Anexo 11: Certificación Summary.

Lic. Jaime Samaniego.

PROFESOR DEL IDIOMA INGLES DEL COLEGIO “8 DE DICEIMBRE” DE LA PAROQUIA SAN PEDRO DE LA BENDITA, CANTÓN CATAMAYO, PROVINCIA DE LOJA.

CERTIFICA:

Que el documento aquí compuesto es fiel traducción del idioma español al idioma ingles del resumen para el trabajo de titulación denominado: **“COMPONENTES DE INTELIGENCIA DE NEGOCIOS EN EL ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES”**, del Sr. EDGAR IVÁN TUZA CUENCA, egresado de la carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja.

Lo certifico en honor a la verdad y autorizo al interesado hacer uso del presente en lo que a sus intereses convenga.

San Pedro de la Bendita, Catamayo, Loja, 25 de abril de 2015.



Lic. Jaime Samaniego.

PROFESOR DEL IDIOMA INGLES DEL COLEGIO “8 DE DICEIMBRE”



Anexo 12: Certificación Revisión Literaria.

Lic. Carolina Paccha Cuenca.

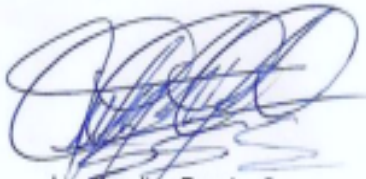
DOCENTE DE LENGUA Y LITERATURA DEL COLEGIO “8 DE DICEIMBRE” DE LA PAROQUIA SAN PEDRO DE LA BENDITA, CANTÓN CATAMAYO, PROVINCIA DE LOJA.

CERTIFICA:

Que el egresado: Edgar Iván Tuza Cuenca, autor del trabajo titulado: **“COMPONENTES DE INTELIGENCIA DE NEGOCIOS EN EL ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES”** que en la redacción e interpretación del contenido de su trabajo ha cumplido con todas las normas y reglas ortográficas, las mismas que han sido revisadas minuciosamente.

Es todo cuanto puedo certificar en honor a la verdad, pudiendo los interesados hacer uso del presente en lo que estime conveniente.

San Pedro de la Bendita, Catamayo, Loja, 25 de abril de 2015.



Lic. Carolina Paccha Cuenca.

DOCENTE DE LENGUA Y LITERATURA DEL COLEGIO “8 DE DICEIMBRE”



Anexo 13: Artículo Científico.