



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

**TRABAJO DE GRADO PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERA EN SISTEMAS COMPUTACIONALES**

TEMA:

**DATAMART PARA EL ANÁLISIS DE INFORMACIÓN DEL SISTEMA
ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE CON
HERRAMIENTAS DE SOFTWARE LIBRE**

AUTORA:

Gloria Estefanía Tana Paspuel

DIRECTOR:

Ing. José Luis Rodríguez

Ibarra – Ecuador

2014



UNIVERSIDAD TÉCNICA DEL NORTE

BIBLIOTECA UNIVERSITARIA

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

1. IDENTIFICACIÓN DE LA OBRA

La UNIVERSIDAD TÉCNICA DEL NORTE dentro del proyecto Repositorio Digital Institucional determina la necesidad de disponer los textos completos de forma digital con la finalidad de apoyar los procesos de investigación, docencia y extensión de la Universidad.

Por medio del presente documento dejo sentada mi voluntad de participar en este proyecto, para lo cual ponemos a disposición la siguiente investigación:

DATOS DEL CONTACTO	
CEDULA DE IDENTIDAD	0401539937
APELLIDOS Y NOMBRES	Tana Paspuel Gloria Estefanía
DIRECCIÓN	Fray Vacas Galindo 4-27 y Mariano Acosta
EMAIL	tana_estefania@yahoo.es
TELÉFONO FIJO	-
TELÉFONO MÓVIL	0988460128
DATOS DE LA OBRA	
TÍTULO	“DATAMART PARA EL ANÁLISIS DE INFORMACIÓN DEL SISTEMA ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE CON HERRAMIENTAS DE SOFTWARE LIBRE”
FECHA	02 de Diciembre del 2014
PROGRAMA	Pregrado
TÍTULO POR EL QUE OPTA	Ingeniería en Sistemas Computacionales
DIRECTOR	Ing. José Luis Rodríguez

2. AUTORIZACIÓN DE USO A FAVOR DE LA UNIVERSIDAD

Yo, Gloria Estefanía Tana Paspuel, con cedula de ciudadanía N° 0401539937, en calidad de autora y titular de los derechos patrimoniales de la obra o trabajo de grado descrito anteriormente, hago la entrega del ejemplar respectivo en formato digital y autorizo a la Universidad Técnica del Norte, la publicación de la obra en el Repositorio Digital Institucional y el uso del archivo digital en la Biblioteca de la Universidad con fines académicos, para ampliar la disponibilidad del material y como apoyo a la educación, investigación y extensión, en concordancia con la Ley de Educación Superior artículo 144.

Ibarra, a los 02 días del mes de Diciembre del 2014

.....


Nombre: Estefanía Tana

C. I.: 0401539937

.....


Ing. Betty Chávez

Jefe de Biblioteca



UNIVERSIDAD TÉCNICA DEL NORTE

CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE INVESTIGACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

Yo, Gloria Estefanía Tana Paspuel, con cedula de ciudadanía N° 0401539937, manifiesto mi voluntad de ceder a la Universidad Técnica del Norte los derechos patrimoniales consagrados en la ley de propiedad intelectual del Ecuador, artículo 4, 5 y 6, en calidad de autora del trabajo de grado denominado: **"DATAMART PARA EL ANÁLISIS DE INFORMACIÓN DEL SISTEMA ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE CON HERRAMIENTAS DE SOFTWARE LIBRE"**, ha sido desarrollado para optar por el título de Ingeniería en Sistemas Computacionales, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En mi condición de autora, me reservo los derechos morales de la obra antes mencionada, aclarando que el trabajo aquí descrito es de mi autoría y que no ha sido previamente presentado para ningún grado o calificación profesional.

En concordancia suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la biblioteca de la Universidad Técnica del Norte.

Firma

Nombre: Gloria Estefanía Tana Paspuel

Cedula: 0401539937

Ibarra, a los 02 días del mes de Diciembre del 2014

CERTIFICACIÓN

Certifico que el trabajo de grado titulado: "DATAMART PARA EL ANÁLISIS DE INFORMACIÓN DEL SISTEMA ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE CON HERRAMIENTAS DE SOFTWARE LIBRE", ha sido realizado en su totalidad por la egresada Gloria Estefanía Tana Paspuel portadora de la cedula de ciudadanía N°. 0401539937.

A handwritten signature in blue ink, appearing to read "José Luis Rodríguez", is written over a horizontal line. The signature is stylized and cursive.

Ing. José Luis Rodríguez

DIRECTOR DE TESIS



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CONSTANCIAS

Yo, Gloria Estefanía Tana Paspuel, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo los derechos de propiedad intelectual correspondiente a este trabajo, a la Universidad Técnica del Norte, según lo establecido por la Ley de Propiedad Intelectual por su reglamento y por la normatividad institucional vigente.

Ibarra, a los 02 días del mes de Diciembre del 2014.

LA AUTORA:

Gloria Estefanía Tana Paspuel

C.I.: 0401539937

DEDICATORIA

El desarrollo del presente proyecto está dedicado a las personas que directa o indirectamente aportaron en la culminación del presente proyecto de tesis.

A mis padres por su apoyo incondicional a lo largo de mi vida, quienes aportaron su confianza en la culminación de esta nueva etapa, llena de sacrificios y obstáculos.

Estefanía Tana

AGRADECIMIENTO

A Dios por haber guiado mi camino durante este largo trayecto estudiantil, quien me ha brindado fuerza en las adversidades y fortaleza para continuar cada día.

A la Facultad de Ingeniería en Ciencias Aplicadas de la Universidad Técnica del Norte por los conocimientos impartidos.

A mi familia, que de alguna manera me apoyaron, brindándome su confianza y cariño en cada reto por cumplir, motivándome a seguir con la finalización de mis objetivos propuestos.

A los docentes de la Facultad de Ingeniería en Ciencias Aplicadas por los conocimientos impartidos, al Ing. José Luis Rodríguez, por su apoyo en la dirección del desarrollo del presente proyecto de tesis.

Estefanía Tana



UNIVERSIDAD TÉCNICA DEL NORTE

Ibarra-Ecuador

UNIVERSIDAD TÉCNICA DEL NORTE Facultad de Ingeniería en Ciencias Aplicadas DECANATO

MEMORANDO 221

PARA: Distinguidos Compañeros (as) FICA
DE: Decano de la Facultad de Ingeniería en Ciencias Aplicadas
ASUNTO: Facilidades
FECHA: 7 de febrero de 2012

Con la finalidad de que la señorita Gloria Estefanía Tana Paspuel, con cédula de ciudadanía 0401539937, estudiante de la Carrera de Ingeniería en Sistemas Computacionales desarrolle la tesis en la Facultad, cuyo tema es: "DATAMART PARA EL ANÁLISIS DE INFORMACIÓN DEL SISTEMA ACADÉMICO DE LA UNIVERSIDAD TÉCNICA DEL NORTE CON HERRAMIENTAS DE SOFTWARE LIBRE"; agradeceré brindar las facilidades del caso.

Atentamente,
CIENCIA Y TÉCNICA AL SERVICIO DEL PUEBLO

Ing. Edwin Rosero Rosero MSc.



Anexo: 5 anillados

Elaborado por:
Pamela Méndez Rhea
2012-02-07

ÍNDICE DE CONTENIDOS

CONTENIDOS	PÁG.
Caratula	
Autorización de Uso y Publicación a favor de la Universidad Técnica del Norte	ii
Cesión de Derechos de Autor del Trabajo de Investigación a favor de la Universidad Técnica del Norte	iv
Certificación del Director de Tesis	v
Constancias	vi
Dedicatoria	vii
Agradecimiento	viii
Certificación de auspicio	ix
Índice de contenidos	x
Índice de figuras	xiii
Índice de tablas	xvi
Resumen	xvii
Summary	xviii
CAPÍTULO I	
INTRODUCCIÓN	
1.1 Descripción de la organización	1
1.1.1 Proceso de Gestión Académica	1
1.1.2 Descripción de procedimientos en la Facultad de Ingeniería en Ciencias Aplicadas	2
1.1.3 Carreras que integran la Facultad de Ingeniería en Ciencias Aplicadas	5
1.1.4 Diagnostico preliminar	5
1.1.5 Requerimientos solicitados	6
1.2 Definición del problema	7
1.2.1 Objetivo general	7
1.2.2 Objetivos específicos	7
1.2.3 Límites	7
1.2.4 Justificación	8
CAPÍTULO II	
MARCO TEÓRICO	
2.1 Introducción	10
2.2 Inteligencia de Negocios	11
2.2.1 Definición de Inteligencia de Negocios	12
2.2.2 Beneficios de la Inteligencia de Negocios	12
2.3 Datawarehousing	13
2.4 ¿Qué es un Datawarehouse?	13
2.4.1 Motivos para desarrollar un datawarehouse	15
2.5 Términos asociados con datawarehouse	15
2.5.1 Datamining	15
2.5.2 OLAP	17
2.5.2.1 Ventajas de OLAP	17

2.5.3	ROLAP	18
2.5.3.1	Ventajas de ROLAP	18
2.5.3.2	Desventajas de ROLAP	19
2.5.4	MOLAP	20
2.5.4.1	Ventajas de MOLAP	20
2.5.4.2	Desventajas de MOLAP	21
2.5.5	Hybrid OLAP	21
2.5.5.1	Ventajas de HOLAP	22
2.5.6	OLTP	22
2.5.6.1	Ventajas de OLTP	23
2.6	Arquitectura de un datawarehouse	24
2.6.1	Datos operacionales o fuentes de datos	24
2.6.2	ETL	24
2.6.3	Datawarehouse	25
2.6.4	Datamart	25
2.6.5	Nivel de acceso a los datos	25
2.6.6	Customer Relationship Management	25
2.6.7	ERP	26
2.7	Costos de un datawarehouse	28
2.7.1	Costos de construcción	28
2.7.2	Costos de operación	29
2.8	Impactos de un datawarehouse	29
2.8.1	Impactos humanos	29
2.8.2	Impactos empresariales	30
2.8.3	Impactos organizacionales	31
2.8.4	Impactos técnicos de datawarehouse	32
2.9	Consideraciones de análisis al implementar un datawarehouse	33
2.10	Metodología para la implementación de un datawarehouse	34
2.10.1	Metodología de Ralph Kimball	34
2.10.1.1	Ciclo de vida del modelo dimensional	36
2.10.2	Metodología de Bill Inmon	38
2.11	Diseño de un datawarehouse usando herramientas de Software Libre	40

CAPÍTULO III

CONSTRUCCIÓN DE LA SOLUCIÓN

3.1	Herramientas de Software Libre	41
3.1.1	Birt	41
3.1.2	Attivio	43
3.1.2.1	Características de Active Intelligence Engine	45
3.1.3	Pentaho	48
3.1.3.1	Descripción	49
3.1.3.2	Productos de Pentaho	49
3.2	Comparativa de herramientas	50
3.2.1	Etapas que se deben seguir para la comparativa de herramientas	51
3.2.2	Resultados	51
3.2.3	Cuadro comparativo de herramientas de software libre	54
3.3	Selección de la herramienta adecuada	55
3.4	Selección del proceso	55

3.5 Granularidad de la información	56
3.5.1 Determinación de la granularidad de la tabla de hechos	57
CAPÍTULO IV	
DISEÑO PROTOTIPO	
4.1 Identificación de las dimensiones para el diseño del prototipo	58
4.2 Identificación de los hechos o métricas	59
4.3 Implementación	59
4.3.1 Diseño del Prototipo	60
4.3.2 Herramientas de Software Libre que se utilizarán	60
4.3.3 Modelamiento de datos lógico	66
4.3.4 Construcción del Prototipo	84
4.3.5 Desarrollo de dimensiones a través de una interfaz gráfica	84
4.4 Puesta en marcha	127
4.5 Pruebas	128
4.6 Resultados de las pruebas obtenidas	129
4.7 Manual de usuario	129
CAPÍTULO V	
CONCLUSIONES Y RECOMENDACIONES	
5.1 Conclusiones	130
5.2 Recomendaciones	132
Glosario de términos	133
Bibliografía	139
ANEXOS	143

ÍNDICE DE FIGURAS

CONTENIDOS	PÁG.
CAPÍTULO I	
Figura 1.1. Diagrama de la creación del prototipo	8
CAPÍTULO II	
Figura 2.1. Niveles de funcionamiento de Business Intelligence implementado en una empresa	12
Figura 2.2. Arquitectura de un datawarehouse empresarial	27
Figura 2.3. Ejemplo de Modelado Estrella propuesta por Ralph Kimball	35
Figura 2.4. Modelamiento Copo de Nieve propuesto por Ralph Kimball	35
Figura 2.5. Modelamiento propuesto por Bill Inmon	39
CAPÍTULO IV	
Figura 4.1. Logotipo de la Suite Pentaho	60
Figura 4.2. Logotipo de la Suite Pentaho Reporting	61
Figura 4.3. Logotipo de la Suite Pentaho Data Integration	62
Figura 4.4. Herramienta Spoon para generar ETL	63
Figura 4.5. Diagrama origen de datos	71
Figura 4.6. Dimensión DIM_ALUMNO	71
Figura 4.7. Dimensión DIM_PROFESOR	72
Figura 4.8. Dimensión DIM_MATERIA_CURSO	72
Figura 4.9. Dimensión DIM_CUSO_PARALELO	72
Figura 4.10. Dimensión DIM_CARRERA_MATERIA	72
Figura 4.11. Dimensión DIM_TIEMPO	73
Figura 4.12. Base de datos dimensional alumno_materia	82
Figura 4.13. Base de datos dimensional carrera_materia	83
Figura 4.14. Insertando un Table Input	84
Figura 4.15. Configuración de un table Input	85
Figura 4.16. Configuración de conexión de base de datos de un Table Input	85
Figura 4.17. Probando conexión de base de datos	86
Figura 4.18. Visualización de datos cargados en un Table Input	87
Figura 4.19. Agregando una secuencia	87
Figura 4.20. Agregando una nueva conexión con la base de datos dimensional	88
Figura 4.21. Probando conexión con la base de datos dimensional	89
Figura 4.22. Configuración de secuencia de una nueva dimensión	89
Figura 4.23. Insertando Get System Info	90
Figura 4.24. Configuración de Get System Info	90
Figura 4.25. Insertando un Insert/Update	91
Figura 4.26. Configuración herramienta de Inserción/Actualización	91
Figura 4.27. Agregando una nueva conexión con la base de datos dimensional	92
Figura 4.28. Probando conexión con la base de datos dimensional	93
Figura 4.29. Configuración de claves primarias	94
Figura 4.30. Configuración de campos de actualización	94
Figura 4.31. Ventana de Mapping de campos de relación	95

Figura 4.32. Configuración de Mapping de campos de relación	95
Figura 4.33. Configuración de Mappings	96
Figura 4.34. Sentencias SQL que ejecuta la herramienta Insertar / Actualizar internamente	96
Figura 4.35. Agregando una secuencia para errores	97
Figura 4.36. Configuración de secuencia de una nueva dimensión	97
Figura 4.37. Unión de las herramientas Insertar/Actualizar con secuencia	98
Figura 4.38. Configuración del Error Handling Step	98
Figura 4.39. Agregando un Table Output	99
Figura 4.40. Configuración del Table Output	100
Figura 4.41. Agregando una nueva conexión con la base de datos dimensional	101
Figura 4.42. Probando conexión con la base de datos dimensional	101
Figura 4.43. Configurando los campos del Database Fields	102
Figura 4.44. Ventana de Mapping de campos de relación	102
Figura 4.45. Configuración de campos de relación para la tabla de auditoría	103
Figura 4.46. Configuración de Mappings	103
Figura 4.47. Sentencias SQL que ejecuta la herramienta Table Output internamente	104
Figura 4.48. Finalización de configuración de una dimensión completa	104
Figura 4.49. Visualización de datos cargados en la FACT	105
Figura 4.50. Insertando un Search in Database	106
Figura 4.51. Configuración de un Search in Database	106
Figura 4.52. Configuración de claves en un Search in Database	107
Figura 4.53. Configuración de valores a devolver en un Search in Database	107
Figura 4.54. Insertando un Memory Group By	107
Figura 4.55. Configuración de campos en un Memory Group By	108
Figura 4.56. Configuración de la FACT_DW	109
Figura 4.57. Configuración de la FACT_DATOS_DW	109
Figura 4.58. Creación del nuevo Cubo Alumno_Materia	110
Figura 4.59. Creación del nuevo Cubo Carrera_Materia	110
Figura 4.60. Creación de dimensiones y jerarquías en el nuevo Cubo	111
Figura 4.61. Cubo completo Alumno_Materia	112
Figura 4.62. Cubo completo Cubo_Carrera	112
Figura 4.63. Configuración de usuario para publicación del Cubo	113
Figura 4.64. Selección de ubicación para publicación del Cubo	113
Figura 4.65. Publicación del Cubo	114
Figura 4.66. Página principal de la consola de administración	115
Figura 4.67. Configuración con la base de datos	115
Figura 4.68. Página inicial del Servidor Mondrian	116
Figura 4.69. Menú principal de Pentaho Bi Server	116
Figura 4.70. Creación de una nueva vista de análisis	117
Figura 4.71. Vista del cubo de información Alumno_Materia	117
Figura 4.72. Vista del cubo de información Carrera_Materia	118
Figura 4.73. Instalación del plugin Saiku en la consola de administración	118
Figura 4.74. Vista del cubo con Saiku en la consola de administración	119

Figura 4.75. Vista del informe generado Total alumnos	121
Figura 4.76. Vista del informe generado alumnos con relación a la materia	122
Figura 4.77. Listado de carreras con sus respectivas Materias	123
Figura 4.78. Listado de carreras	126

ÍNDICE DE TABLAS

CONTENIDOS	PÁG.
CAPÍTULO I	
Tabla 1.1. Descripción de procedimientos académicos en la FICA	4
CAPÍTULO III	
Tabla 3.1. Comparativa de herramientas	54
CAPÍTULO IV	
Tabla 4.1. Herramientas de software libre para desarrollar un datawarehouse	60
Tabla 4.2. Definición de parámetros para tablas y secuencias	67
Tabla 4.3. Definición de parámetros para campos de dimensiones	68
Tabla 4.4. Definición de parámetros para secuencias	68
Tabla 4.5. Definición de parámetros para el proceso ETL	69
Tabla 4.6. Mapeo de datos de origen	74
Tabla 4.7. Mapeo de información de la dimensión DIM_ALUMNO	75
Tabla 4.8. Mapeo de información de la dimensión DIM_PROFESOR	76
Tabla 4.9. Mapeo de información de la dimensión DIM_MATERIA_CURSO	77
Tabla 4.10. Mapeo de información de la dimensión DIM_CURSO_PARALELO	78
Tabla 4.11. Mapeo de información de la dimensión DIM_CARRERA_MATERIA	79
Tabla 4.12. Mapeo de información de la dimensión DIM_TIEMPO	81

RESUMEN

La inteligencia de negocios es un proceso que permite realizar la toma de decisiones en las distintas áreas del negocio, esta integración de datos permite convertirlos en información valiosa con la finalidad de que el usuario final sea capaz de determinar oportunidades y amenazas para la empresa a través de los reportes generados; para que la implementación tenga éxito se requiere de datos históricos o datos que provengan de distintas fuentes de datos que, consolidados, organizados y clasificados conforman un datamart y la reunión de varios de estos conforman un datawarehouse, que es una base de datos independiente para realizar consultas y análisis de información de manera rápida y confiable al usuario. Actualmente existen muchas soluciones de software que permiten realizar inteligencia de negocios ya sean de software libre o software propietario, la decisión por optar alguna de estas soluciones la tiene el personal encargado del departamento de tecnologías de información junto con su equipo de desarrollo, ellos deben de considerar algunos factores para la implementación de una solución de Inteligencia de Negocios como son: los recursos con los que cuenta la empresa, la predisposición que tienen los altos mandos en el apoyo del proyecto, la cooperación de todo el personal que conforma la organización y la disponibilidad de información requerida para ser analizada. La implementación de un prototipo de datamart para el análisis de información se lo hizo con la finalidad de demostrar que el uso de herramientas de software libre en la actualidad está en auge en las pequeñas y medianas empresas, el prototipo cumplió con diferentes etapas que fueron realizadas con las distintas herramientas que ofrece la Suite de Pentaho, la información del datamart está almacenada en una base de datos PostgreSQL que también es de software libre.

SUMMARY

Business intelligence is a process that allows decision making in different business areas, this integration allows converting data into valuable information in order that the end user is able to identify opportunities and threats for the company through reports generated; for successful implementation requires historical data or data derived from different data sources, consolidated, organized and classified form a datamart and meeting several of these make up a datawarehouse, which is a separate database for querying and analyzing information quickly and reliably to the user. Currently there are many software solutions that enable business intelligence whether free software or proprietary software, the decision to choose one of these solutions lies with the personnel department of information technology along with its development team, they must consider some factors in implementing a Business Intelligence solution as they are: the resources available to the company, the predisposition of high command in supporting the project, cooperation of all staff that makes up the organization and availability of information required to be analyzed. The implementation of a prototype datamart for information analysis did in order to demonstrate that the use of free software tools today is booming in small and medium enterprises, the prototype comply with different stages were performed with the various tools offered by Pentaho Suite, datamart information is stored in a PostgreSQL database which is also free software.

CAPÍTULO I

INTRODUCCIÓN

1.1 DESCRIPCIÓN DE LA ORGANIZACIÓN

Se explica en forma general la descripción de los directivos de la Facultad y las respectivas las funciones de que cumplen cada uno de ellos.

1.1.1 Proceso de Gestión Académica

El proceso de gestión académica que se realiza en las distintas carreras que integran la Facultad de Ingeniería en Ciencias Aplicadas (FICA), dirigida por diferentes docentes de la facultad, suelen presentar distinta complejidad según las tareas que le sean asignadas, la facultad tiene docentes designados en diferentes cargos como son: Decanato, Subdecanato, Directores de escuela de las distintas carreras, Coordinadores de carrera quienes son los encargados de dirigir y supervisar el bienestar académico de los estudiantes; esto aporta complejidades adicionales al trabajo académico ya que los resultados son supervisados por entes superiores como son el Rectorado, Vicerrectorado Académico en la orientación, control, medición y evaluación de los resultados académicos.

Para profundizar la información sobre los procesos académicos se define a todos aquellos trámites que el estudiante realiza durante su carrera académica como son los procesos de matrícula, pagos, evaluación de docentes, consulta de notas y horarios, defensa de anteproyecto, defensa de tesis, etc.; un proceso de registro académico inicia con la admisión del estudiante a la Universidad y se lleva un registro durante su vida académica hasta su graduación, el objetivo de llevar un registro es tener ordenada, ágil y de manera oportuna la información de toda la actividad académica desarrollada por cada uno de los estudiantes universitarios, pero también hay que tomar en cuenta que las autoridades que dirigen la facultad realizan actividades que complementan los procesos académicos del estudiante como es la aprobación de anteproyectos, aprobación de notas de defensas privadas de los trabajos de grado, aprobaciones de anulaciones de matrículas, etc.

1.1.2 Descripción de procedimientos en la Facultad de Ingeniería en Ciencias Aplicadas

Nº	Área Involucrada	Personal Responsable	Actividades Desarrolladas
1	Estudiante	Estudiante	<ul style="list-style-type: none"> • Ingreso del estudiante a la Facultad. • Consulta de notas, horarios de clase. • Defensa de anteproyecto. • Defensa de trabajo de grado.
2	Secretaría	Secretaria	<ul style="list-style-type: none"> • Atención al estudiante. • Recepción de papeles en la matrícula. • Matriculación del estudiante. • Legalización de matrícula. • Emitir reportes de calificaciones.
3	Dirección de Escuela	Director de Escuela	<ul style="list-style-type: none"> • El control de asistencia de docentes es de absoluta responsabilidad del Director de Escuela conjuntamente con el Secretario Abogado (Honorable Consejo Universitario, 2005). • El control del avance programático de las asignaturas, es de estricta responsabilidad del Director de Escuela (Honorable Consejo Universitario, 2005). • Cumplir con todas las disposiciones emanadas por las Autoridades de la Facultad (Honorable Consejo Universitario, 2005).
4	Sub-Decanato	Subdecano	<ul style="list-style-type: none"> • Planificar y organizar las jornadas curriculares (Honorable Consejo Universitario, 2005). • Planificar y organizar las jornadas curriculares (Honorable Consejo Universitario, 2005). • La gestión académica de la Facultad es de responsabilidad del Subdecano (Honorable Consejo Universitario, 2005).

Nº	Área Involucrada	Personal Responsable	Actividades Desarrolladas
4	Sub-Decanato	Subdecano	<ul style="list-style-type: none"> • Presentar al Decano los informes de labores cumplidas y la planificación del siguiente periodo académico (Honorable Consejo Universitario, 2005). • Coordinar la organización y planificación de la investigación, ciclos de conferencias, paneles, talleres, seminarios, cursos y otras actividades académicas (Honorable Consejo Universitario, 2005). • Presentar al Decano los informes de labores cumplidas y la planificación del siguiente periodo académico (Honorable Consejo Universitario, 2005). • Coordinar la organización y planificación de la investigación, ciclos de conferencias, paneles, talleres, seminarios, cursos y otras actividades académicas (Honorable Consejo Universitario, 2005). • Elaborar y presentar el Calendario Académico Interno de la Facultad al H. Consejo Directivo para su aprobación (Honorable Consejo Universitario, 2005). • Presidir y coordinar los concursos de merecimientos y oposición para la provisión de profesores titulares de la Facultad (Honorable Consejo Universitario, 2005). • Presentar proyectos de mejoramiento académico referentes al proceso de enseñanza – aprendizaje (Honorable Consejo Universitario, 2005). • Presidir la Comisión de Escalafón Docente de la Facultad y actuar de conformidad con el reglamento pertinente (Honorable Consejo Universitario, 2005).

Nº	Área Involuc.	Personal Responsa.	Actividades Desarrolladas
5	Decanato	Decano	<ul style="list-style-type: none"> • Responsable de planificar, dirigir, supervisar y ejecutar, las actividades académicas, administrativas y financieras de la Facultad (Honorable Consejo Universitario, 2005). • Es el encargado de la gestión ejecutiva de las actividades de la Facultad (Honorable Consejo Universitario, 2005). • Es responsable de la administración de los Fondos de Autogestión y Autofinanciamiento e informará de sus resultados operativos al H. Consejo Directivo (Honorable Consejo Universitario, 2005). • El Decano previa autorización del H. Consejo Directivo, solicitará la adquisición de bienes para equipamiento de la Facultad y prestación de servicios, de conformidad con los ordenadores de gastos y pago vigente y solicitará al Rector la suscripción de los respectivos contratos con fondos provenientes de la autogestión y autofinanciamiento (Honorable Consejo Universitario, 2005).
6	Honorable Consejo Directivo	Integrantes Honorable Consejo Directivo	<ul style="list-style-type: none"> • Encargados de aprobar o denegar las solicitudes hechas por los estudiantes (Honorable Consejo Universitario, 2005). • Presentar al Vicerrectorado Académico las actividades planificadas y realizadas en la Facultad, hasta 15 días antes de inicio del nuevo período académico (Honorable Consejo Universitario, 2005). • Tienen la obligación de presentar en forma oportuna el informe de actividades planificadas y realizadas al Vicerrector Académico para el trámite respectivo en la Comisión Académica (Honorable Consejo Universitario, 2005).

Tabla 1.1: Descripción de procedimientos académicos en la FICA

Fuente: [<http://www.utn.edu.ec/web/portal/images/doc-utn/reglamento-general-utn.pdf>]

1.1.3 Carreras que integran la Facultad de Ingeniería en Ciencias Aplicadas

La Facultad de Ingeniería en Ciencias Aplicadas (FICA) está integrada por las siguientes carreras:

- Carrera de Ingeniería Textil,
- Carrera de Ingeniería en Sistemas Computacionales,
- Carrera de Ingeniería en Electrónica y Redes de Comunicación,
- Carrera de Ingeniería en Mecatrónica,
- Carrera de Ingeniería Industrial.

1.1.4 Diagnóstico preliminar

Es muy importante tomar en cuenta que las pequeñas y medianas empresas están evolucionando de acuerdo al paso agigantado de la tecnología, haciendo uso de servidores para almacenar su información, actualmente las grandes empresas albergan su información en la web con la finalidad de mantener sus datos protegidos contra desastres naturales, pérdida de información por falla de equipos físicos, mala manipulación de la información, etc.

El datawarehouse se ha convertido actualmente, en el centro de atención de las pequeñas y medianas empresas, brinda facilidades para que los usuarios finales exploten de mejor manera la información que puede ser administrada por las distintas aplicaciones tecnológicas, creadas para el análisis de información; al mismo tiempo se define como una colección de datos clasificada e integrada, información utilizada como soporte importante para el proceso en la toma de decisiones gerenciales. Adentrándose más en esta arquitectura se encuentra a los datamarts que es la unión de varios cubos que contienen información, son creados con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Lo ventajoso de esta herramienta es que los datos pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

1.1.5 Requerimientos solicitados

En la actualidad las empresas poco a poco están emigrando todos sus sistemas a herramientas libres para no tener que comprar licenciamiento y así satisfacer sus necesidades actuales, lo que se pretende al realizar este estudio de varias herramientas de software libre es brindar opciones que permitan la creación de datamarts para que las empresas o instituciones interesadas en migrar su información a herramientas software libre sepan cuál es su mejor opción así como las ventajas y desventajas de cada una de estas herramientas.

Cada día la tecnología evoluciona significativamente, es por ello que las pequeñas o medianas empresas migran su información a grandes almacenes de datos, para que al momento de tomar decisiones gerenciales el usuario final haga un uso más óptimo de los datos y así los altos mandos puedan tomar decisiones mensuales o semestrales, según el caso sobre el movimiento de sus negocios y así poder crear operaciones del negocio, pero también se debe considerar las arquitecturas tecnológicas óptimas para la implementación funcional de un datawarehouse.

Las herramientas más conocidas son con licenciamiento de las casas productoras de Microsoft y Oracle que son costosas para las pequeñas y medianas empresas que están en auge de desarrollo, es por ello que se hace necesaria la investigación de herramientas de software libre para contrastar con las herramientas de software propietario para así poder determinar ventajas y desventajas.

En la FICA se tiene vigente un sistema de datos académico desde 1998 hasta la presente fecha, pero no se ha hecho un análisis de la información para obtener tendencias y generar estadísticas para la toma de decisiones en el proceso de gestión académica en cada una de las facultades y hacer un seguimiento más profundo en cuanto al rendimiento y aprovechamiento de los alumnos que integran a la facultad.

1.2 DEFINICIÓN DEL PROBLEMA

Existe poco conocimiento sobre herramientas de software libre para la creación de datamarts por lo que se hace necesario investigar alternativas más económicas y eficientes para realizar inteligencia de negocios, conocer sus ventajas y limitaciones para solucionar problemas de tipo real.

1.2.1 Objetivo general

Estudiar e implementar un Datamart para el análisis de información del Sistema Académico de la Universidad Técnica del Norte con herramientas de software libre.

1.2.2 Objetivos específicos

- Investigar las herramientas libres Birt, Attivio y Pentaho que permiten la creación de datamarts.
- Seleccionar entre las herramientas libres propuestas con la finalidad de determinar la herramienta más adecuada.
- Aplicar la investigación en la realización de un prototipo de datamart con la información del sistema Académico de la FICA.
- Presentar informes y reportes generados a través de los cubos de información.

1.2.3 Límites

Este prototipo es creado exclusivamente para la generación de reportes para el análisis de la información de la Facultad de Ingeniería en Ciencias Aplicadas, permitiendo al usuario tener un panorama claro de la información que generan los estudiantes en las distintas carreras cada semestre generando y llevando la información de manera ordenada y segura en forma automática a través de una base de datos y las herramientas de software libre para la generación de reportes.

Este prototipo no se integrará a la base de datos Oracle existente actualmente en la Universidad Técnica del Norte sino que se trabajará con una base de datos propia generada en PostgreSQL, se utilizarán datos de periodos anteriores a la fecha, para la demostración de la presente investigación.

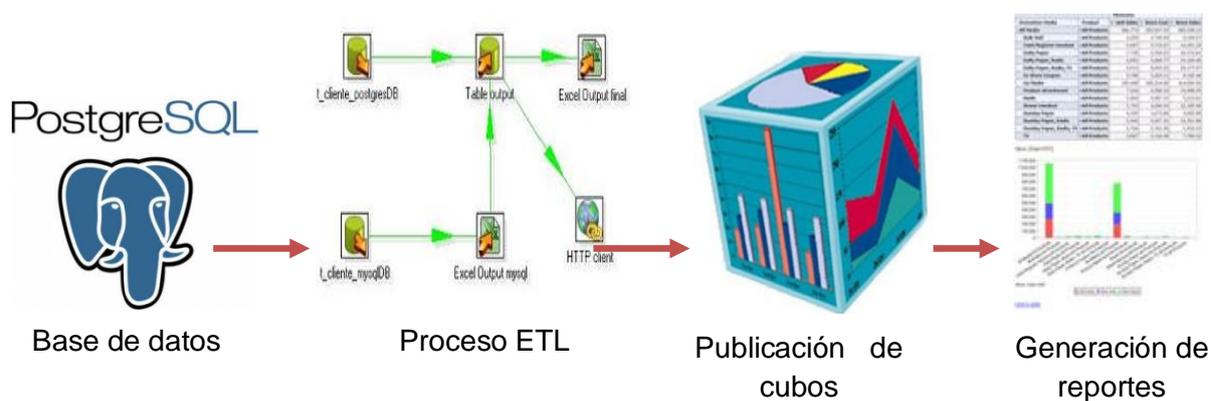


Figura 1.1. Diagrama de la creación del prototipo
Fuente: Elaboración propia

En la figura se encuentra esquematizado el funcionamiento propuesto para la implementación del prototipo a crearse para demostrar la presente investigación; el prototipo pretende demostrar el análisis de información mediante la generación de reportes.

1.2.4 Justificación

La presente investigación justifica la importancia de realizar una investigación de varias herramientas de software libre para la creación de datamarts; es elemental analizar nuevas tecnologías o plataformas libres para que las empresas pequeñas y medianas eviten comprar licenciamiento. La inteligencia de negocios es un tema actual que beneficia al desarrollo económico y social de las pequeñas, medianas y grandes empresas, una acertada decisión hace que los negocios tomen un rumbo próspero y productivo pero más beneficioso les resultaría no tener que pagar licenciamiento por el software que están utilizando.

La inteligencia de negocios se aplica con un enfoque, integracionista y disciplinado en el cual se puede identificar los requerimientos de toma de decisiones operacionales, tácticos y estratégicos y así poder plantear soluciones viables para la empresa o negocio que conjuntamente con bases de datos libres se puedan integrar y trabajar.

La ventaja de las herramientas de software libre es que proporcionan libertad a los usuarios finales sobre su producto pudiendo ser usado, copiado, estudiado, cambiado y redistribuido libremente, entre las ventajas más importantes tenemos:

- Disminuye el costo por el uso de licencias.
- Libertad de uso y redistribución, asegurando independencia tecnológica.
- Soporte y compatibilidad mediante el acceso a foros y blogs.
- Formatos estándar, sistema en expansión, etc.

El uso de software libre reduce el costo de implementación, permitiendo a las pequeñas y medianas empresas implementar una solución de inteligencia de negocios a bajo precio.

CAPÍTULO II

MARCO TEÓRICO

1.1 INTRODUCCIÓN

Actualmente las pequeñas y medianas empresas necesitan depositar su confianza, en la toma de decisiones que el personal encargado de analizar la información reporta sobre sus negocios, y para que dichas decisiones sean lo más reales posible es necesario tener hechos y cifras, se puede evidenciar que la competencia tecnológica crece a pasos agigantados, es por ello que las organizaciones buscan lugares acertados y seguros donde albergar su información; el gran reto de cada organización es saber qué hacer cuando se tiene una montaña de información la misma que necesita ser analizada, organizada y clasificada, lógicamente que el personal encargado de tener en orden esta información necesitará mucho más tiempo de lo estimado, pero con el actual desarrollo de nuevas herramientas de business intelligence para analizar la información es una realidad de presentar información en muy poco tiempo.

Por ese motivo, se está dando auge a las herramientas de software libre que permitan minimizar el tiempo de respuesta para analizar grandes volúmenes de información con la mayor velocidad y precisión posible; al hacer uso de herramientas con licenciamiento de bajo costo se está logrando que las organizaciones mantengan su rendimiento competitivo y a la orden del día. Hay muchos factores que influyen en el desarrollo de las pequeñas y medianas empresas, como son: la globalización del mercado, la presión de la competencia, los avances tecnológicos, el crecimiento de las microempresas, etc., esto hace que las organizaciones se vuelvan más competitivas en la rama de inteligencia de negocios esta es una clara muestra por la que las empresas invierten en tecnología y soluciones informáticas, las empresas no dependen únicamente de factores geográficos sino también de conocimientos en tecnología que es muy importante para mantener una ventaja competitiva.

La inteligencia de negocios es el resultado de un proceso minucioso que hace uso de la información a través de herramientas ETL, éste proceso que luego de haber recolectado, procesado, limpiado y transformado la información necesaria para la toma de decisiones, pasa a ser una tarea no tan sencilla si consideramos que una empresa tiene distintas áreas donde su información necesita ser analizada y centralizada en un solo bloque de información.

1.2 INTELIGENCIA DE NEGOCIOS

La inteligencia de negocios es la habilidad para transformar los datos en información entendible, capaz de permitir a los usuarios sacar conclusiones sobre sus movimientos económicos, financieros, transaccionales, etc., con la finalidad de optimizar al máximo el proceso de toma de decisiones en las distintas áreas del negocio, éste proceso juega un papel estratégico muy importante para una empresa u organización en su proceso de desarrollo, generando una potencial ventaja competitiva como es la inserción a nuevos mercados, lanzamiento de promociones u ofertas de productos, eliminación de información redundante e innecesaria, control financiero, reajuste de costos, planificación del lote total de la producción, análisis de perfiles que pueden ser de clientes, usuarios, altos mandos, empleados, rentabilidad en la venta de un producto nuevo, análisis del perfil y rendimiento estudiantil, etc. **(Sinnexus, 2014)**

Los principales productos de business intelligence que existen hoy en día son:

- Cuadros de Mando Integrales.
- Sistemas de Soporte a la Decisión.
- Sistemas de Información Ejecutiva.
- Minería de datos.
- Datamart
- Datawarehouse

2.2.1. Definición de Inteligencia de Negocios

“La inteligencia de negocios es el conjunto de productos y servicios que permiten a los usuarios acceder y analizar de manera rápida y sencilla la información generada en la empresa para la toma de decisiones del negocio a nivel operativo, táctico y estratégico”. (Deinsa, 2014)



Figura 2.1 Niveles de funcionamiento de business intelligence implementado en una empresa
Fuente: [<http://www.idensa.com/>]

En el año 2000, la clave para surgir en el mercado laboral competitivo era mantenerse delante de sus competidores, es por ello que las empresas grandes empezaron a utilizar aplicaciones de inteligencia de negocios con licenciamiento, corriendo el riesgo de elevar el costo en sus inversiones, éstas aplicaciones deben estar continuamente disponibles a nivel de empleados, consultores, clientes, proveedores y el público en general, y si el costo de licenciamiento es menor es una ventaja que nos permite para tomar decisiones correctas basadas en información exacta y actualizada.

2.2.2. Beneficios de la Inteligencia de Negocios

Las empresas pueden caer en errores comunes como tener mucha información disponible y no saber qué hacer con ella, por este motivo que se hace necesario hacer uso de la inteligencia de negocios para la solución a éste problema, a continuación se describen los beneficios más relevantes:

- Mayor facilidad para analizar estrategias de precios.

- Mejorar la toma de decisiones de forma más rápida, informada y basada en la información real generada por la empresa.
- Mejorar el acceso a los datos a través de consultas, análisis o reporte de informes, con información actualizada.
- Control adecuado del flujo de información.
- Mayor integración de la información.
- Aumentar el control de costos y gastos.
- Analizar la productividad de los empleados.

1.3 DATAWAREHOUSING

Ante las frecuentes dificultades presentadas por los sistemas tradicionales en satisfacer las necesidades de información a los gerentes, surge el datawarehouse, como solución a las necesidades de información globales de la empresa, este es conocido comúnmente como almacén de datos.

La ventaja primordial de tener implementado un datawarehouse se basa principalmente en la estructuración de la información, este proceso garantiza un almacenamiento homogéneo y confiable, en una estructura basada en la consulta y el tratamiento jerarquizado de las fuentes de datos.

1.4 ¿QUÉ ES UN DATAWAREHOUSE?

Datawarehouse término que traducido al español significa **almacén de datos**, podremos definirlo como una colección de datos integrado, orientado a un determinado campo de acción que puede ser una empresa, una organización, una institución educativa, etc., su objetivo es ayudar a la toma de decisiones en el lugar donde esté implementado. Estos almacenes de datos están integrados de grandes volúmenes de información que se subdividen en unidades lógicas más pequeñas denominadas datamarts dependiendo de la cantidad de información que maneje la entidad. **(Wikipedia, 2014)**

Un Datawarehouse es una base de datos dimensional que es la encargada de integrar y depurar toda la información de una o más bases de datos transaccionales distintas, para luego procesar la información permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta; al implementar un datawarehouse en pequeñas o medianas empresas representa el primer paso para implantar una solución completa y segura de inteligencia de negocios.

La finalidad de construir e implementar un datawarehouse en una empresa es poder analizar la información histórica que posee la empresa, analizando los movimientos y transacciones en tiempos pasados y en el presente, con el objetivo de decidir las futuras acciones y toma de decisiones en el presente.

Entre las principales ventajas al implementar un datawarehouse son:

- Un datawarehouse es una herramienta que permite tomar decisiones en cualquier área funcional que se basa en información integrada, organizada y global del área del negocio.
- El datawarehouse proporciona a la aplicación técnicas para el análisis y modelización de la información con la finalidad de permitir la generación de reportes a través de la base de datos dimensional.
- Es de gran ayuda dentro de la empresa u organización en la implantación de sistemas de gestión integral con el usuario final.

A continuación se describe el concepto de datawarehouse de dos expertos en la materia:

Bill Inmon para diseñar un datawarehouse: “utiliza una metodología descendente ya que de esta manera se considerarán a todos los datos corporativos”. **(Wikipedia, 2014)**

Ralph Kimball, otro especialista en datawarehouse lo define como: “una copia de las transacciones de datos estructurada para la consulta y el análisis”. **(Wikipedia, 2014)**

Otro punto importante es que Kimball también definió a un datawarehouse como la unión de varios Datamarts implementados y generados en una entidad; la metodología que utiliza para diseñar un datawarehouse es la ascendente.

2.4.1 Motivos para desarrollar un datawarehouse

A continuación se detalla los motivos por los cuales una empresa u organización debe implementar un datawarehouse:

- Es muy importante la implementación de un datawarehouse para realizar tareas en los servidores asociados a reportes y queries¹ que no son utilizados por sistemas de proceso de transacciones.
- Proporcionar información clave para la toma de decisiones empresariales.
- Los datawarehouse son una herramienta muy útil para el almacenamiento, consulta y análisis de datos históricos.
- Para proveer a los servidores de aplicaciones una información ordenada para que pueda ser explotada y reportada de la mejor manera.
- Las empresas utilizan modelos de datos o tecnologías avanzadas en los servidores que agilizan los queries y reportes.

2.5 TÉRMINOS ASOCIADOS CON DATAWAREHOUSE

2.5.1 Datamining

El datamining es el estudio y tratamiento de datos masivos para extraer conclusiones e información relevante de ellos, datamining que en español significa “minería de datos”, se encuentra ubicado en el nivel estratégico que posee la empresa, es decir, es la información clasificada que tiene la alta gerencia para la toma de decisiones su objetivo es eliminar o minimizar las fallas ocasionadas por el talento humano al analizar la información y dejar que sean los datos los que se muestren en los reportes para la toma de decisiones gerencial, la minería de datos implica el uso de técnicas de bases de datos como los índices espaciales. **(Wikipedia, 2014)**

¹ **Queries:** Es una consulta o una búsqueda de datos almacenados en una base de datos. **Fuente:** Tomado de Diccionario de Informática, Definición de Query [<http://www.alegsa.com.ar/Dic/query.php>]

La tarea primordial de la minería de datos es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones desconocidos, como:

- Los grupos de registros de datos (análisis de clúster),
- Registros poco usuales (detección de anomalías) y
- Dependencias (reglas de la minería de datos).

El datamining utiliza algoritmos y técnicas avanzadas de análisis cuantitativo e inteligencia artificial para revisar los datos, pudiendo ser utilizado en cualquier situación o proceso que genere información. Algunas de las aplicaciones donde se aplica la minería de datos son las siguientes: **(Wikipedia, 2014)**

2.5.1.1 Optimización de Campañas de Marketing: Se utiliza datamining para determinar que potenciales clientes tienen una mayor probabilidad de responder a una campaña específica, pudiendo así reducir costos con la información exacta para la mejor toma de decisiones.

2.5.1.2 Retención de Clientes: Ayuda a la compañía a reaccionar oportunamente a identificar aquellos clientes que tienen una alta probabilidad de fuga.

2.5.1.3 Riesgo de Crédito: La minería de datos sirve para crear un modelo de evaluación de crédito de deudores, identificando las características particulares de un cliente para determinar si es o no riesgoso.

2.5.1.4 Logística y Control de Inventario: Ayuda a predecir niveles de stock, optimiza canales de distribución, etc.

Como conclusión se puede afirmar que la minería de datos es el proceso de analizar la información desde diferentes perspectivas con el objetivo de resumir los datos en segmentos de información útiles para luego tomar la decisión más acertada, esto se lo puede hacer mediante los dashboards que son parte de la inteligencia de negocios.

2.5.2 OLAP

On-Line Analytical Processing, que traducido significa “Procesamiento analítico en línea”, éste tipo de procesamiento es una estrategia utilizada en el campo de la inteligencia negocios, cuya finalidad es reducir el tiempo de respuesta ante una consulta en grandes cantidades de datos. Para cumplir con este objetivo utiliza estructuras multidimensionales o los muy conocidos cubos OLAP los mismos que se componen de medidas clasificadas en dimensiones; éstos contienen datos seleccionados y clasificados de las grandes bases de datos transaccionales, son creados a partir de un esquema en estrella o copo de nieve. **(Wikipedia, OLAP, 2014)**

OLAP es muy utilizado en la generación de reportes desde distintas áreas como pueden ser: ventas, marketing, dirección gerencial, talento humano, minería de datos y áreas similares, una de las ventajas más importantes es el tiempo de respuesta en las consultas; si analizamos una base de datos relacional podemos determinar que almacena entidades en tablas discretas siempre y cuando hayan sido normalizadas previamente, esta estructura es eficiente en un sistema OLTP pero para OLAP las consultas complejas multi-tabla son evidentemente lentas.

2.5.2.1 Ventajas de OLAP

- El acceso a los datos lo realiza a través del método de sólo lectura, esto es común en las consultas con pocas inserciones, actualizaciones o eliminaciones.
- Los datos se estructuran y organizan de acuerdo a las áreas del negocio y la información está integrada de manera uniforme en toda la base de datos.
- El historial de datos que se guarda es a largo plazo, puede ser de dos a cinco años.
- Las bases de datos OLAP se alimentan de información procedente de los sistemas operacionales existentes, mediante el proceso ETL.

- Además el procesamiento en línea es ventajoso porque ejecuta de manera rápida y eficiente sentencias SQL de tipo SELECT, lo que no ocurre con OLTP que es una buena opción para operaciones de tipo INSERT, UPDATE Y DELETE.

Las aplicaciones OLAP son utilizadas por analistas y gerentes que frecuentemente requieren una vista de datos de nivel superior, hay casos en los que se requiere reportes de las ventas totales por línea de producto, por región, por ciudad, por provincia, etc., las bases de datos OLAP normalmente se actualizan en lote y muy frecuentemente de múltiples fuentes, estas proporcionan un back-end² analítico poderoso a las aplicaciones de múltiples usuarios, es por ello que las bases de datos OLAP se perfeccionan para el análisis de datos en la empresa u organización.

A continuación se describe los diferentes tipos de sistemas OLAP, cada uno tiene ciertos beneficios y desventajas al momento de ser implementado en cada una de las empresas u organizaciones:

2.5.3 ROLAP

Sus siglas significan: “Procesamiento Analítico en línea Relacional”, es una implementación OLAP que almacena los datos en un motor relacional, es decir que los datos son detallados para evitar que las tablas se encuentren sin normalizar, a ésta arquitectura la compone un servidor de banco de datos relacional y un motor OLAP implementado en un servidor para realizar esta función, la ventaja primordial de esta arquitectura es, que nos permite el análisis de grandes cantidades de información. Éste procesamiento analítico permite acceder a la información de una base de datos relacional por medio de consultas SQL que permitirá calcular la información al nivel apropiado cuando el usuario final lo requiera. **(Wikipedia, ROLAP, 2013)**

2.5.3.1 Ventajas de ROLAP

² **Back-end:** Trabaja conjuntamente con el front-end que es el responsable de recolectar los datos de entrada del usuario, que pueden ser de muchas y variadas formas, y los transforma ajustándolos a las especificaciones que demanda el back-end para poder procesarlos, devolviendo generalmente una respuesta que el front-end recibe y expone al usuario de una forma entendible para este. La conexión del front-end y el back-end es un tipo de interfaz. **Fuente:** Tomado de Wikipedia, Front-end y back-end [http://es.wikipedia.org/wiki/Front-end_y_back-end]

Las herramientas OLAP que son implementadas sobre base de datos relacionales, tienen ventajas frente a base de datos multidimensionales:

- El procesamiento analítico en línea relacional es más escalable, permitiendo manejar y analizar grandes cantidades de información.
- Permite re-ajustar el proceso inicial de ETL a un modelo de datos específico con la finalidad de que los tiempos de demora en la carga de los datos sean mucho menores.
- Ésta herramienta permite que los datos sean almacenados en bases de datos relacionales, permitiendo que la información pueda ser leída desde cualquier herramienta que permite la generación de informes SQL.
- Es posible modelar datos con éxito que de otro modo no se ajustarían en un modelo dimensional estricto.

2.5.3.2 Desventajas de ROLAP

- Las herramientas ROLAP tienen menor rendimiento que las herramientas MOLAP.
- Para las herramientas ROLAP el proceso ETL debe ser gestionado personalmente, esto implica que se necesita más tiempo en el desarrollo e implementación.
- Es muy frecuente observar que en algunos modelos dimensionales ROLAP no se crean tablas agregadas, es por ello que se afecta el tiempo de rendimiento de una consulta y el acceso a las tablas de información debe ser más detallada.
- Las herramientas ROLAP no poseen algunas de las funcionalidades especiales de las herramientas MOLAP, que puede ser por ejemplo el indexado jerárquico especial de los datos.
- Las nuevas versiones de ROLAP vienen con mejoras en el lenguaje SQL, como operadores CUBE y ROLLUP, vistas de cubo DB2 y extensiones SQL OLAP.

- Las herramientas ROLAP no son tan aconsejables cuando se realiza muchas operaciones de cálculo que no podrían traducirse de manera eficiente en SQL.

2.5.4 MOLAP

El procesamiento analítico multidimensional en línea es una implementación OLAP que permite almacenar información en una base de datos multidimensional, con el objeto de minimizar los tiempos de respuesta cada vez que se accede a la base de datos, muestra información resumida que es usualmente pre-calculada. Éste sistema multidimensional en línea está compuesto de una arquitectura de dos niveles: una base de datos multidimensional que se encarga del manejo, acceso y obtención la información y un motor analítico; el nivel de aplicación es el encargado de ejecutar los requerimientos OLAP, el nivel de presentación se integra con el nivel de aplicación y proporciona una interfaz a través de la cual los usuarios finales pueden visualizar los análisis OLAP. **(Wikipedia, MOLAP, 2014)**

Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional, ya que la información procedente de los sistemas operacionales se cargan en el sistema MOLAP mediante una serie de rutinas por lotes; una vez cargado el dato elemental en la Base de Datos se realiza una serie de cálculos por lotes; luego de rellenar esta estructura, se generan unos índices y algoritmos de tablas Hash para mejorar los tiempos de accesos a las consultas.

2.5.4.1 Ventajas de MOLAP

- Optimización en el almacenamiento de la información, esto hace que las consultas se ejecuten de manera óptima debido a la ventaja de indexación multidimensional.
- Posee técnicas de compresión en los datos, lo que permite ocupar menor espacio en disco.

- Esta herramienta permite la automatización en el procesamiento de los datos agregados de mayor nivel, es muy compacta para conjuntos de datos de pocas dimensiones.
- El modelo de almacenamiento en vectores o matrices proporciona una indexación natural.
- Posee una eficaz extracción de datos gracias a la pre-estructuración que realiza en los datos agregados.

2.5.4.2 Desventajas de MOLAP

- Para grandes volúmenes de datos la carga puede ser bastante extensa, la solución sería utilizar un procesamiento incremental, es decir que solo se debería hacer el procesamiento de los datos que han sido modificados en lugar de volver a procesar todo el conjunto de datos.
- Las herramientas MOLAP tienen muchas dificultades al momento de realizar las consultas de los modelos con dimensiones muy altas.
- MOLAP tiene muchas dificultades al momento de realizar la actualización y consulta de los modelos con más de diez dimensiones.
- El enfoque MOLAP introduce redundancia en los datos.

2.5.5 Hybrid OLAP

Hybrid Online Analytical Process que significa: procesamiento analítico en línea híbrido que es un desarrollo reciente con la combinación de las arquitecturas ROLAP y MOLAP para brindar una solución eficaz con las características de desempeño superior y gran escalabilidad, HOLAP permite almacenar una parte de los datos como en un sistema MOLAP y el resto como en un ROLAP, lo que quiere decir que el almacenamiento de una parte de datos lo hace en un motor relacional y la otra parte en una base de datos multidimensional. **(Wikipedia, HOLAP, 2013)**

HOLAP utiliza dos técnicas importantes:

- **Particionamiento vertical.-** HOLAP almacena agregaciones para mejorar la velocidad de las consultas como un MOLAP y las consultas se detallan en ROLAP para optimizar el tiempo en que se procesa el cubo.
- **Particionamiento horizontal.-** HOLAP almacena una sección de datos normalmente los más recientes en modo MOLAP para mejorar la velocidad de las consultas, y los datos más antiguos en ROLAP, se pueden almacenar algunos cubos en MOLAP y otros en ROLAP.

HOLAP es usado para cubos que requieren rápida respuesta, para sumalizaciones³ basadas en una gran cantidad de datos.

2.5.5.1 Ventajas de HOLAP:

- Mayor rendimiento que la herramienta MOLAP.
- Permite navegar a través de las capacidades de ROLAP.
- Al utilizar ambas arquitecturas automáticamente se implementa el drilling⁴ el cual genera sentencias SQL para recuperar registros detallados para mayor análisis, esto le brinda al programador la percepción de que está pasando de la base de datos multidimensional a la relacional.
- Los cubos almacenados con HOLAP, son más pequeños que los MOLAP y responden con mayor rapidez en la generación de resultados.

2.5.6 OLTP

Es un procesamiento que facilita la administración de operaciones transaccionales, facilita la recuperación y procesamiento de transacciones mediante un gestor transaccional; ésta herramienta está basada en una arquitectura cliente-servidor, ya que es muy utilizada por empresas con una red informática distribuida. **(Wikipedia, OLTP, 2013)**

³ **Sumarización:** Muestra los datos de una manera más resumida, permitiendo, precisamente, calcular valores agregados, que no son los datos directos registrados, sino datos derivados de ellos. **Fuente:** Tomado de Sumarización [http://sumarizacionyessengmiranda.blogspot.com/2010/12/definicion.html]

⁴ **Drilling:** Significa navegación entre arquitecturas que puede ser horizontal o vertical.

OLTP sus siglas traducidas significan: “Procesamiento de Transacciones en Línea”, se utiliza en innumerables aplicaciones como en procesamiento de pedidos, comercio electrónico, supermercados, industria y en banca electrónica, este sistema responde de inmediato a las peticiones del usuario, un ejemplo claro es un cajero automático de un banco ya que es una aplicación de procesamiento de transacciones comerciales. **(Wikipedia, OLTP, 2013)**

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones, de manera sencilla definiremos a esta transacción que se genera como un proceso atómico que se valida como un commit⁵, o invalidado con un rollback⁶ y que puede involucrar operaciones de inserción, modificación y borrado de datos, este proceso transaccional es muy básico en las bases de datos operacionales.

Actualmente, los sistemas transaccionales han presentado dos innovaciones como son: la optimización de las bases de datos OLTP para el procesamiento de transacciones y la especialización de bases de datos analíticas para el procesamiento analítico.

2.5.6.1 Ventajas de OLTP

- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura, por ejemplo la enorme cantidad de transacciones que tienen que soportar las bases de datos.
- Los datos se estructuran según el nivel de aplicación.
- Los formatos de los datos no son necesariamente uniformes, en cada uno de los diferentes departamentos.
- El historial de datos suele limitarse a los datos actuales o recientes.

⁵ **Commit:** Sentencia en SQL que finaliza una transacción de base de datos dentro de un sistema gestor de base de datos relacional (RDBMS) y pone visibles todos los cambios a otros usuarios. **Fuente:** Tomado de Wikipedia, Commit [<http://es.wikipedia.org/wiki/Commit>]

⁶ **Rollback:** Operación que devuelve a la base de datos a algún estado previo. Los Rollbacks son importantes para la integridad de la base de datos, la base de datos puede ser restaurada a una copia limpia incluso después de que se han realizado operaciones erróneas. **Fuente:** Tomado de Wikipedia, RollBack [<http://es.wikipedia.org/wiki/Rollback>]

- Las empresas están trasladando los datos transaccionales al Cloud Computing⁷, así que las aplicaciones rentables basadas en transacciones pasan de “comprar” a “alquilar”, es por ello que las empresas modernas ahora son empresas híbridas que procesan datos tanto dentro de la empresa como en el Cloud Computing o en la nube.
- Las aplicaciones OLTP se caracterizan por la creación de muchos usuarios, actualizaciones o recuperación de registros individuales, por lo tanto las bases de datos OLTP se perfeccionan para actualización de transacciones.

Actualmente uno de los problemas que los desarrolladores intentan resolver con la tecnología relacional son los datos multidimensionales por ejemplo, una consulta SQL para crear resúmenes de ventas de un producto podrían involucrar la revisión de la mayoría, o de todos los registros en una base de datos de mercadeo y podría tomar horas de proceso, es por ello que se hace necesario el uso de un servidor OLAP que podría ocuparse de todos estos procesos en unos segundos.

Las aplicaciones OLTP generalmente no requieren de datos históricos, por lo tanto las aplicaciones OLTP y las bases de datos tienden a ser organizadas; es decir “orientadas al tema” alrededor de procesos específicos, respondiendo a muchos requerimientos de las altas gerencias.

2.6 ARQUITECTURA DE UN DATAWAREHOUSE

Un datawarehouse tiene una arquitectura compleja para ser analizada, esta es una forma de representar la estructura global de los datos, la comunicación, los procesos y la presentación al usuario final, es muy importante describir los procesos y pasos internos que se crean y complementan para que los cubos tengan éxito. La arquitectura de un datawarehouse está constituida por las siguientes partes:

⁷ **Cloud computing:** La computación en la nube son servidores desde Internet encargados de atender las peticiones en cualquier momento. Se puede tener acceso a su información o servicio, mediante una conexión a internet desde cualquier dispositivo móvil o fijo ubicado en cualquier lugar. **Fuente:** Tomado de Wikipedia, Computación en la Nube [http://es.wikipedia.org/wiki/Computaci%C3%B3n_en_la_nube]

- 2.6.1 Datos operacionales o fuentes de datos.**- Es el origen de datos, son las diferentes bases de datos en los que está almacenada la información aquí se incluyen los ERPs, CRMs, archivos planos, bases de datos como ORACLE, MYSQL, ACCESS, POSTGRESQL, SQL SERVER, etc., es el componente de almacenamiento físico del datawarehouse.
- 2.6.2 ETL.**- Que significa: **extracción de datos**, que es la selección sistemática de información usada para formar parte del datawarehouse; **transformación de datos**, son los procesos para sumarizar y realizar cambios en los datos operacionales; **carga de datos**, que es la inserción de datos de diferentes fuentes. Éste proceso hace que la información se ingrese de manera ordenada y sistemática desde múltiples fuentes, éstos datos pueden ser reformateados, limpiados y cargados en una nueva base de datos dimensional; el objetivo del proceso ETL es que los datos primarios luego de la transformación, validación y carga realice un proceso de filtración interna para poner a disposición esta información y posteriormente ser analizada por los usuarios.
- 2.6.3 Datawarehouse.**- Almacenamiento físico de datos, es el gran almacén de datos donde se almacena toda la información resumida, los metadatos y los datos atómicos que son los datos de las tablas normalizadas de las fuentes de datos.
- 2.6.4 Datamart.**- Los datamarts son almacenes de datos más pequeños que por lo general son utilizados por las pequeñas y medianas empresas a diferencia de los cubos de información que nos ayudan a visualizar la información clasificada.
- 2.6.5 Nivel de acceso a los datos.**- Son herramientas que proveen acceso a la información estandarizada, a estos datos tienen acceso los usuarios, administradores de base de datos, se establece niveles de acceso a la información con la finalidad de protegerla, por su grado de confidencialidad en todas las empresas, se puede definir los niveles de acceso de acuerdo a los roles que desempeñen en la empresa con un usuario y una contraseña o una clave de acceso.

2.6.6 Customer Relationship Management.- Conocido como CRM es una estrategia de negocio focalizada hacia el cliente, como a todas las aplicaciones informáticas, tanto de software como de hardware, necesarias para procesar, analizar y exponer la información resultante, con la finalidad de medir y retroalimentar la estrategia de negocio desarrollada, mediante el análisis de las informaciones extraídas por los clientes desde los diferentes canales o medios de comunicación.

2.6.7 ERP.- Son sistemas de planificación de recursos empresariales que integran y manejan muchos de los negocios asociados con las operaciones de producción y de los aspectos de distribución de una compañía comprometida en la producción de bienes o servicios.

Los objetivos principales de los sistemas ERP son:

- Optimización de procesos empresariales.
- Integridad de datos, acceso a la información de forma confiable, precisa y oportuna.
- Posibilidad de compartir información entre todos los componentes de la organización.
- Eliminación de datos y operaciones innecesarias de reingeniería.

También hay que tomar en cuenta que las diferentes bases de datos también son fuentes de información que se integran al proceso ETL y permiten la generación de los respectivos cubos, datamarts, datawarehouse, para su posterior generación de reportes para la toma de decisiones gerenciales y empresariales.

ARQUITECTURA DE UN DATAWAREHOUSE

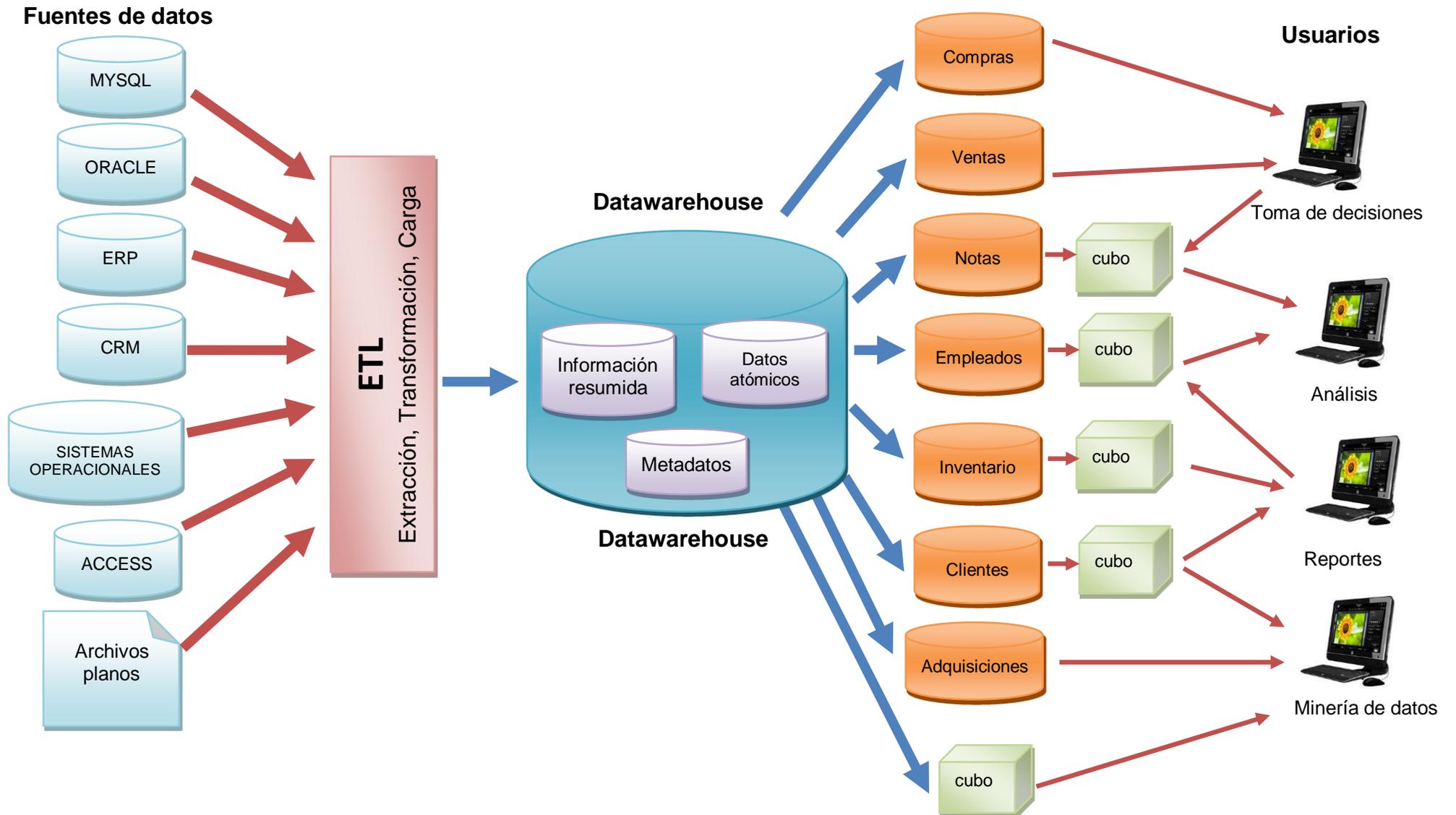


Figura 2.2 Arquitectura de un Datawarehouse empresarial
Fuente: Elaboración propia

2.7 COSTOS DE UN DATAWAREHOUSE

Para considerar los costos de un datawarehouse hay que tomar en cuenta los factores importantes que implica llevar a cabo un proyecto de este tipo, esto depende mucho del tamaño de la empresa y del volumen de información que se maneja en ella, es muy importante saber cómo se ve afectado tanto por los cambios a nivel de OLTP como por el ambiente empresarial, a continuación detallamos los costos más importantes:

2.7.1 Costos de construcción.- Los costos varían como si fuera para cualquier proyecto de tecnología de información, se puede destacar tres categorías a tomar en cuenta:

- **Talento Humano:** los desarrolladores de proyectos deben contar con un conocimiento claro sobre las áreas que maneja la empresa y de los procesos empresariales que se lleva a cabo, también es importante considerar que para el desarrollo de un datawarehouse se requiere la participación de: analistas en la toma de decisiones, programadores, especialistas en el desarrollo de datawarehouse; todas estas personas deben trabajar en equipo, compartiendo sus conocimientos y destrezas para enfrentar los desafíos de desarrollo de un datawarehouse.
- **Tiempo:** Se debe considerar el tiempo que tomará llevar a cabo la elaboración del proyecto, tomando en consideración la construcción y entrega de resultados, también es importante tomar en cuenta el tiempo que conlleva la planeación del proyecto y la elección de la arquitectura adecuada, éstos parámetros establecen un marco de referencia y un conjunto de estándares que son importantes para que el datawarehouse brinde los mejores resultados.
- **Tecnología:** El costo que implica adoptar una nueva tecnología en la empresa puede ser un factor muy importante a la hora de implementar un datawarehouse, y esto se consideraría como la inversión inicial del proyecto.

2.7.2 Costos de operación.- Una vez que está construido y puesto en funcionamiento un datawarehouse, es muy importante tomar en cuenta el soporte y mantenimiento del mismo, se puede distinguir tres tipos de costos de operación:

- **Evolutivos:** ajustes y cambios continuos del datawarehouse a través del tiempo, estos cambios de expectativas pueden resultar por el cambio de personal administrativo o por incremento de áreas en la empresa.
- **Crecimiento:** Pueden haber muchos factores como incremento en el tiempo, incremento en volúmenes de datos, incremento en el número de usuarios del datawarehouse, esto conlleva a un incremento de recursos como: demanda de monitoreo, administración, sintonización del datawarehouse, esto se requiere con la finalidad de evitar un incremento en los tiempos de respuesta y la recuperación de información en el menor tiempo posible.
- **Cambios:** El datawarehouse debe soportar cambios que pueden ocurrir desde los orígenes de sus datos que éste contenga, como en la información almacenada que el datawarehouse soporta.

2.8 IMPACTOS DE UN DATAWAREHOUSE

Para que un datawarehouse tenga impactos positivos en la empresa, su éxito no solo depende de su construcción, sino también en saber cómo explotar la información que genera, para mejorar procesos empresariales y operaciones que se manejan en la empresa y para que un datawarehouse tenga el resultado esperado es necesario entender los impactos de implementación en los siguientes ámbitos:

2.8.1 Impactos Humanos

Se analizarán los efectos sobre el personal que integra la empresa:

2.8.1.1 Construcción del datawarehouse: Construir un datawarehouse requiere la participación activa de todos quienes conforman la empresa y están involucrados en el proyecto, depende de la situación de la empresa, se deben considerar las condiciones tanto tecnológicas, económicas, políticas y sociales, las cuales determinan cómo va a estar estructurado un datawarehouse.

2.8.1.2 Accesando el datawarehouse: El datawarehouse provee datos e información necesaria para su respectivo análisis lo que permite a los usuarios acceder a la información que posee la empresa cuando ellos lo requieran, esta entrega de información implica varias situaciones:

- El personal de la empresa necesita aprender nuevas destrezas y recibir capacitación adecuada.
- Serán eliminados los análisis extensos y las demoras en la obtención de la información. Como esta información estará lista para ser visualizada, las expectativas frente a la eliminación aumentarán significativamente.
- Mejores oportunidades emergen en el campo empresarial para los especialistas en análisis de información.
- La generación de reportes en papel serán reducidos o eliminados ya que los reportes se visualizan de manera digital.
- La robustez del datawarehouse dependerá exclusivamente de su uso activo y la respectiva retroalimentación de información por parte de sus usuarios.

2.8.2 Impactos Empresariales

Procesos y Decisiones Empresariales.- Los impactos empresariales importantes se describen a continuación:

- Las herramientas que ayudan en los procesos de toma de decisiones tienen un rendimiento óptimo mediante la disponibilidad de información, así la toma de decisiones empresariales se hacen más eficientes.
- Los procesos empresariales son óptimos cuando se hayan eliminado los tiempos de respuesta al momento de hacer consultas si la información es incorrecta o no encontrada, la petición es eliminada.
- Las secuencias de procesos empresariales claros y entendibles pueden ser optimizados para ganar eficiencia y reducir costos.
- Al momento de acceder a la información del datawarehouse los datos son examinados de los sistemas operacionales.
- El cliente final aprende mucho de los sistemas de información cuando los datos organizados y estructurados son presentados en reportes ordenados y entendibles.
- Pueden encontrarse algunos posibles defectos en las herramientas actuales, siendo posible entonces mejorar la calidad de nuevas aplicaciones.

2.8.3 Impactos Organizacionales

Cuando el datawarehouse es fuente primaria de información empresarial, los siguientes impactos pueden ser posibles:

- El personal de la empresa debidamente capacitado, que maneja la aplicación tiene mayor confianza en las decisiones empresariales que se toman ya que tienen la certeza de que las decisiones están basadas en información confiable que genera el datawarehouse.
- Es muy importante la delegación de funciones del personal que tiene acceso a la información, de esta manera, cada usuario entiende su propio rol y responsabilidades.
- La información que se visualiza en los reportes es muy entendible lo que permite el rendimiento entre el personal q labora en la empresa, esto

mejora la confianza y cooperación entre distintos sectores que la empresa maneja, viéndose reducida la sectorización de funciones.

- La visibilidad, accesibilidad y conocimiento de la información producen mayor confianza en los sistemas operacionales.

2.8.4 Impactos Técnicos del datawarehouse

Considerando las etapas de construcción y soporte se tiene los siguientes impactos técnicos:

2.8.4.1 Nuevas destrezas de desarrollo: al momento de empezar con la construcción del datawarehouse, el impacto más grande sobre el personal técnico de desarrollo de la empresa está dado por la curva de aprendizaje, muchas destrezas nuevas se deben aprender, como son:

- Conceptos, manejo y estructura de un datawarehouse.
- El datawarehouse introduce muchas tecnologías nuevas como ETL's, Catálogo de Metadatos, base de datos dimensionales, etc., esto cambia la manera de ver en que se usa la tecnología existente.
- Las nuevas responsabilidades de soporte, demandas de recursos y nuevas expectativas, son los efectos de estos nuevos cambios.
- El personal de desarrollo debe mantener habilidades de diseño y muy buen análisis de procesos, es importante tomar en cuenta los requerimientos empresariales que no son posibles de puntualizar de forma estable a través del tiempo.
- Se requiere de técnicas de desarrollo incremental y evolutivo.
- Es importante el trabajo en equipo de todos los colaboradores en el proyecto, personal cooperativo y participantes activos en el desarrollo del proyecto.

2.8.4.2 Nuevas responsabilidades de operación: Los cambios sobre los sistemas y datos operacionales deben ser examinados y analizados con mucho detenimiento para determinar el impacto que estos cambios tienen sobre el datawarehouse.

2.9 CONSIDERACIONES DE ANÁLISIS AL IMPLEMENTAR UN DATAWAREHOUSE

Hay que tomar en cuenta algunos factores al momento de considerar la implementación, implantación y análisis de un datawarehouse ya que este proporciona información de gestión accesible, correcta, uniforme y actualizada a continuación detallamos algunas consideraciones:

- Proporcionar información clave para la toma de decisiones empresariales y gerenciales.
- Mejorar la calidad de las decisiones tomadas.
- Tener los objetivos claros.
- Los sistemas deben ser relativamente sencillos de instalar y manejar.
- Implementar una herramienta muy útil para el almacenamiento de información y consultas de históricos.
- Proporcionar un gran poder de procesamiento de información.
- Permitir mayor flexibilidad y rapidez en el acceso a la información.
- Ayudar a las empresas a obtener un aumento de la productividad.
- Proporcionar una comunicación fiable entre todos los departamentos de la empresa.
- Permitir conocer qué está pasando en el negocio o empresa, con el objetivo de mantener siempre informado al gerente de los buenos y malos resultados.
- Transformar los datos en información y la información en conocimiento

- Reducir los tiempos de respuesta y los costos de operación.
- Proporcionar mayor flexibilidad ante el entorno, mejorar el servicio al cliente y permitir el rediseño de los procesos.

2.10 METODOLOGÍA PARA LA IMPLEMENTACIÓN DE UN DATAWAREHOUSE

Para la implementación de un datawarehouse se debe analizar las opciones que se tiene en metodologías de arquitectura, existen dos metodologías importantes en el desarrollo: la metodología de Kimball y la metodología de Inmon ambas son metodologías de desarrollo.

2.10.1 Metodología de Ralph Kimball

Ralph Kimball propone la metodología de modelamiento dimensional que se rige en el ciclo de vida dimensional del negocio, esta metodología es una de las técnicas más fáciles para construir un datawarehouse, se forma por modelos de tablas y relaciones y en base a las consultas hechas en una base de datos relacional; lo que permite presentar los datos de manera estándar e intuitiva. Los elementos de estas tablas se pueden definir de la siguiente manera:

- **Tabla de hechos:** las tablas de hechos contienen las claves primarias de las tablas dimensionales.
- **Dimensiones:** son las tablas que contienen los campos seleccionados de la base de datos relacional.
- **Medidas:** son los aditivos, cálculos numéricos de una tabla de hechos que representan el comportamiento del negocio a una dimensión.

Cada entrada a la tabla de hechos está conectada a una dimensión lo que permite determinar el contexto de los hechos. Una base de datos dimensional puede conformar un cubo con tres o cuatro dimensiones (OLAP), se puede representar un modelo dimensional a una tabla de hechos rodeada por las tablas de dimensiones, a éste modelamiento se le denomina modelo estrella-uniión.

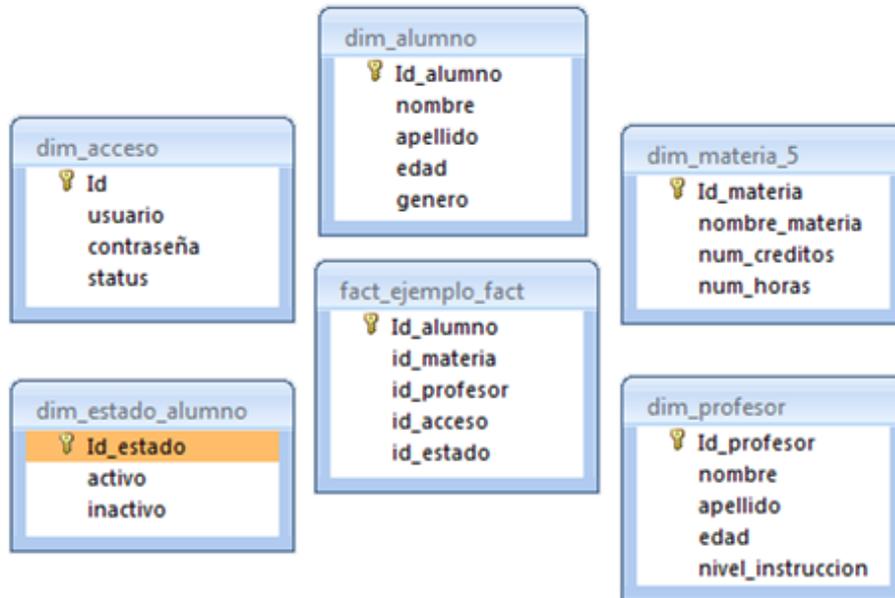


Figura 2.3 Ejemplo de Modelado Estrella propuesta por Ralph Kimball
Fuente: Elaboración propia

Otra clase de Modelamiento es la que se conoce como copo de nieve, esta consta de ramificaciones a partir de las tablas de dimensiones y no solo a partir de la tabla de hechos.

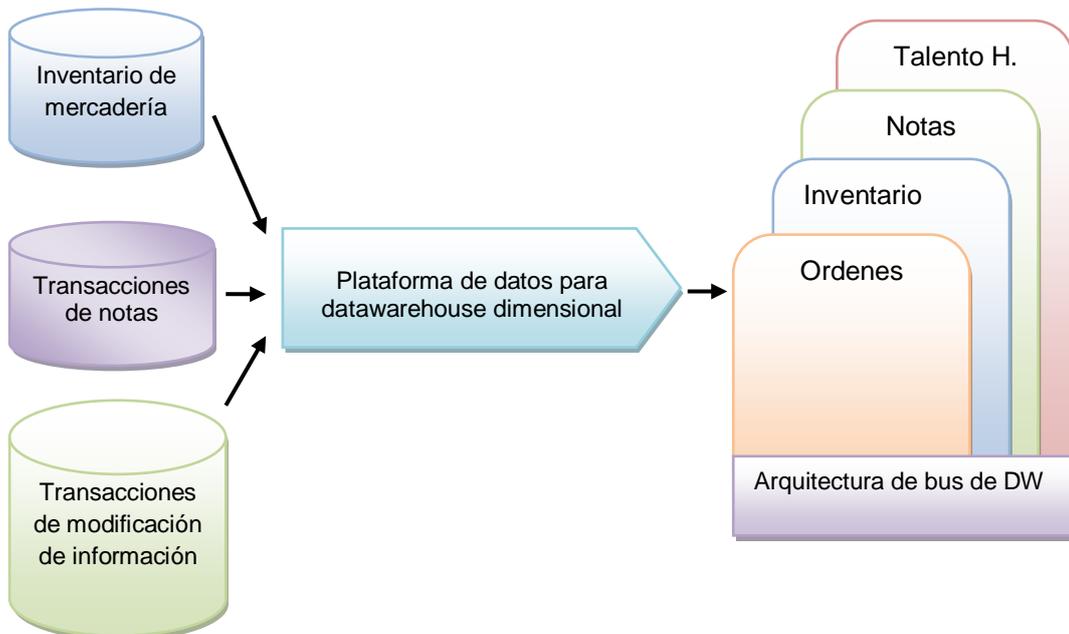


Figura 2.4 Modelamiento Copo de Nieve propuesto por Ralph Kimball
Fuente: Elaboración propia

2.10.1.1 Ciclo de Vida del Modelo Dimensional

La metodología que Ralph Kimball propone las siguientes fases:

- **Planificación del Proyecto:** en esta etapa se define el alcance y sus limitaciones del proyecto de creación de un datawarehouse, se define la cantidad de personal a ocupar en el desarrollo del proyecto, se desarrolla el plan de proyecto su seguimiento, monitorización y capacitación respectiva al personal que va a usar el proyecto.
- **Definición de los Requerimientos:** en esta fase los programadores deben tener definido cuáles son los factores claves del negocio para determinar los requerimientos, levantar estos requerimientos y convertirlos en consideraciones de diseño óptimas para el éxito de un diseño de datawarehouse.
- **Modelado Dimensional:** para empezar a modelar se diseña una matriz donde se determina, la dimensionalidad de cada tabla de la base de datos transaccional para determinar los diferentes grados de detalle dentro de cada concepto del negocio.
- **Diseño Físico:** en esta fase se hace un estudio minucioso y se elige las estructuras necesarias que soporten el diseño lógico, aquí se hace una definición de estándares del entorno de la base de datos, su estructuración, indexación y las estrategias de particionamiento.
- **Diseño y Desarrollo de la presentación de datos:** esta etapa es muy importante, es la base fundamental en el desarrollo del datawarehouse a través de actividades de ETL. En éste proceso se define la materia prima, que son los datos para luego pasar a la siguiente etapa.
- **Diseño de la arquitectura técnica:** en esta fase se deben tener en cuenta tres factores: los requerimientos que tiene la empresa, la arquitectura tecnológica, las directrices técnicas y estratégicas planificadas a futuro por la empresa, esto permitirá establecer un diseño arquitectónico estable del entorno del datawarehouse.

- **Selección de productos e instalación:** aquí el personal encargado del proyecto deberá evaluar y seleccionar los componentes de la arquitectura tecnológica a utilizar como puede ser plataforma de hardware, motor de BD, herramienta ETL's, arquitectura de software, etc.
- **Especificación de aplicaciones para clientes finales:** aquí se definen e identifican los roles o perfiles de los usuarios para los diferentes tipos de aplicaciones o para los distintos campos de aplicación para los cuales va encaminado el proyecto.
- **Desarrollo de aplicaciones para clientes finales:** aquí se involucran configuraciones de los metadatos y la respectiva construcción de reportes específicos de acuerdo al área del negocio.
- **Implementación:** en esta fase se prueba la correcta implementación y funcionamiento de la tecnología, la información y las aplicaciones que los usuarios finales son accesibles para el área del negocio.
- **Mantenimiento y crecimiento:** en esta etapa surge la necesidad de continuar con actualizaciones de forma constante, mantenimientos de las aplicaciones para lograr el constante funcionamiento de las metas que se plantearon al iniciar el proyecto.
- **Gestión del proyecto:** aquí el personal encargado del monitoreo y mantenimiento del proyecto asegura que el ciclo de vida del proyecto se lleve a cabo de manera sincronizada.

Además de aplicar la metodología de Ralph Kimball se debe complementar con actividades de análisis de las áreas de negocio, requerimientos de necesidades con los niveles estratégicos e intermedios de la empresa, esto se lo hace con la finalidad de conocer las diferentes decisiones que toma cada responsable de cada área de negocio. Es necesario tomar en cuenta que para los procesos de carga de información se cuenta con especialistas en desarrollo de procesos de base de datos dimensionales que ayudan a implementar de manera eficiente todo aquel proceso necesario para pasar la información de las fuentes oficiales a la base de datos dimensional del datawarehouse.

Y por último, se debe contar con herramientas conocidas en el mercado para la explotación de información, la experiencia de los consultores en inteligencia de negocios es muy importante, esto asegura a nuestros clientes la calidad total en el desarrollo e implementación del proyecto de inteligencias de negocios.

2.10.2 Metodología de Bill Inmon

La metodología que Bill Inmon plantea es transferir la información de las diferentes herramientas OLTP, que son los Sistemas Transaccionales de las empresas a un lugar organizado, donde la información pueda ser utilizada para el análisis, además describe las siguientes características:

- **Orientado a temas.**- La información contenida en la base de datos transaccional deben estar organizados, es decir, todos los datos relacionados al mismo evento u objeto deben quedar unidos entre sí.
- **Integrado.**- La fuente de información debe contener los datos integrados de todos los módulos operacionales de la organización o empresa, estos datos deben ser consistentes y confiables.
- **No volátil.**- Una ventaja importante es que la información no se debe modificar, ni eliminar, una vez almacenado un dato en la base de datos dimensional se convierte en información de sólo lectura y se conserva para futuras consultas.
- **Variante en el tiempo.**- Los cambios producidos en los datos a lo largo del tiempo quedan registrados, es decir existe un historial de información, para que los informes que se puedan generar en un futuro reflejen esas variaciones.

La información debe estar simplificada al nivel de máximo de detalle. A nivel empresarial los cubos o datamarts departamentales son tratados como subconjuntos de un datawarehouse corporativo, éstos son construidos para cubrir las necesidades individuales de análisis de información de cada departamento.

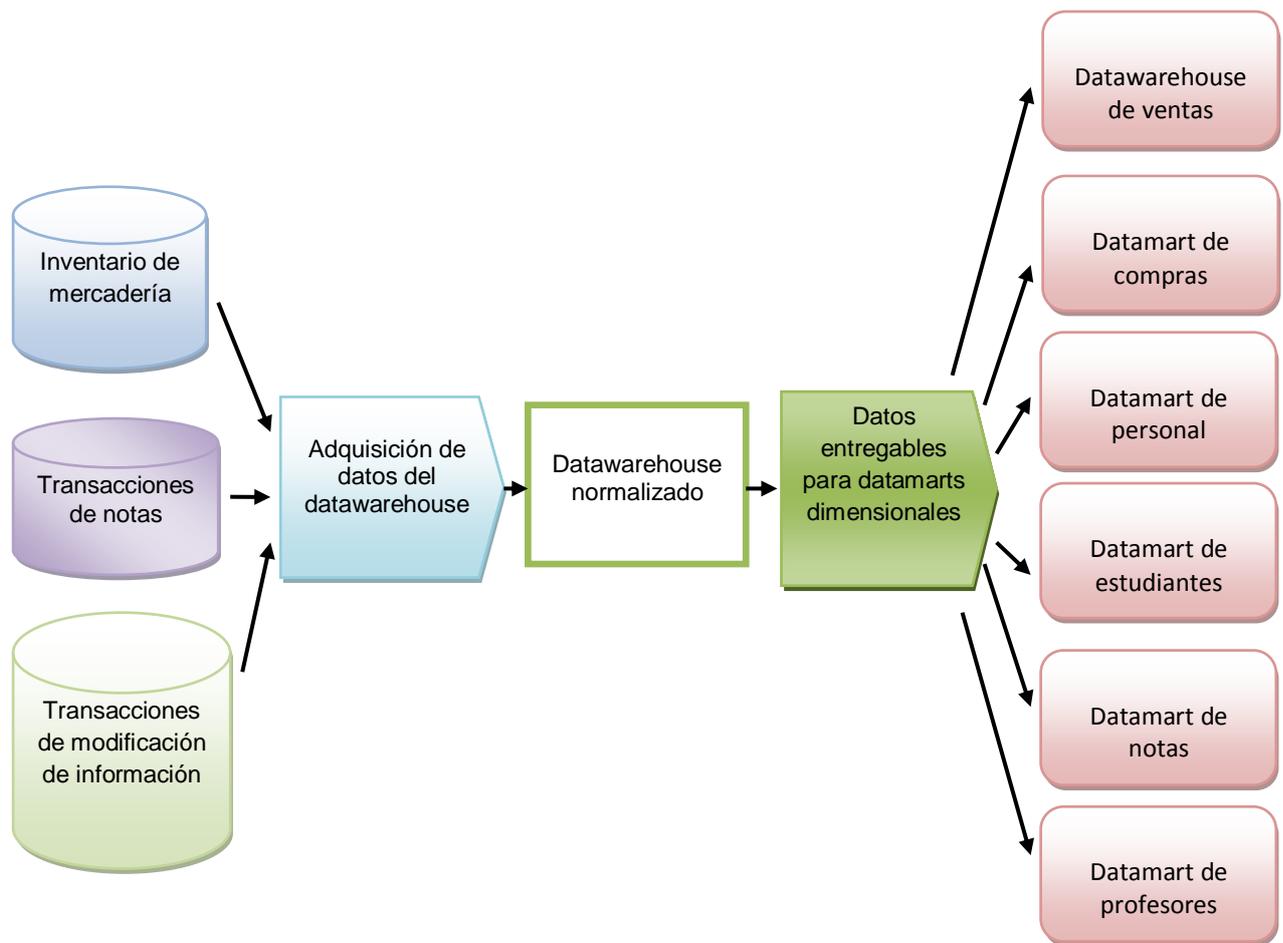


Figura 2.5 Modelamiento propuesto por Bill Inmon
Fuente: Elaboración propia

Al enfoque de la metodología de Inmon también se le conoce como Top-down⁸, esto quiere decir que la información es extraída de los sistemas operacionales a través de los procesos ETL y son debidamente cargados en las áreas de almacenamiento dimensionales, donde son validados, consolidados y organizados en el datawarehouse, donde existen los llamados metadatos que ayudan en la documentación de forma clara y precisa del contenido de la información, una vez realizado el proceso de actualización de los datamarts generados en el proyecto, se obtiene la información con las transformaciones hechas por el programador, se organizan los datos en estructuras particulares, refrescando su contenido.

⁸ **Top-down:** Formula un resumen del sistema, sin especificar detalles, es una estrategia de procesamiento de información.
Fuente: Tomado de Wikipedia, Top-down y bottom-up [http://es.wikipedia.org/wiki/Top-down_y_bottom-up]

La metodología para la construcción de un proyecto de este tipo es un poco difícil para un sistema de información enorme, por otra parte para el tratamiento de los cambios en los datos se usa la metodología de insertar fechas en los datos para determinar su validez.

2.11 DISEÑO DE UN DATAWAREHOUSE USANDO HERRAMIENTAS DE SOFTWARE LIBRE

La solución al utilizar herramientas de software libre en el desarrollo de inteligencia de negocios se ha convertido en un interesante ejercicio de análisis de la arquitectura de desarrollo.

Herramientas como Pentaho, Birt, Alfresco, OpenBravo, Talend, Palo, Eclipse Birt, JasperReports⁹, OpenL, etc., son las que ayudan en el desarrollo de los datamarts y cubos ya que esto forma un proceso centralizado bajo un marco de trabajo orientado a la solución de problemas de soporte de decisión, empleando componentes de inteligencia de negocios y permitiendo desarrollos completos para atender a soluciones de toma de decisiones, estas herramientas permiten el análisis y la presentación de los datos, cabe recalcar que las herramientas ETL no son consideradas generalmente como herramientas de inteligencia de negocios.

La mayor parte de las herramientas de business intelligence son utilizadas por el cliente final para acceder, analizar y reportar datos convertidos en reportes que muy comúnmente residen en un datawarehouse, datamarts y almacenes de datos operacionales.

Una plataforma que integra componentes Open Source¹⁰ mostrando una combinación de flujos de trabajo y administración de procesos es una verdadera herramienta de trabajo en el desarrollo de datamarts.

⁹ **JasperReports:** Es una librería de creación de informes que tiene la habilidad de entregar contenido enriquecido al monitor, a la impresora o a ficheros. **Fuente:** Tomado de Wikipedia, JasperReports [http://es.wikipedia.org/wiki/JasperReports]

¹⁰ **Open Source:** Código abierto es la expresión con la que se conoce al software distribuido y desarrollado libremente. **Fuente:** Tomado de Wikipedia, Código abierto [http://es.wikipedia.org/wiki/C%C3%B3digo_abierto]

CAPÍTULO III

CONSTRUCCIÓN DE LA SOLUCIÓN

3.1 HERRAMIENTAS DE SOFTWARE LIBRE

A continuación se detallará, a cada una de las herramientas de software libre:

3.1.1 BIRT

Es un proyecto de Eclipse de código abierto, el cual ofrece un gran número de tareas, sobre todo para aplicaciones web, con capacidades de inteligencia de negocios para clientes pesados. Birt recibe soporte de una activa comunidad de usuarios y desarrolladores de eclipse.org. **(Wikipedia, 2014)**

Esta herramienta trabaja con dos elementos primordiales:

- un componente de tiempo de ejecución para generar informes que se puedan extender en cualquier entorno de Java.
- el segundo componente, es un motor gráfico que se puede implementar en el diseñador de informes, se puede utilizar para integrar gráficos.

Haciendo un poco de historia el código inicial del proyecto BIRT fue diseñado y desarrollado por Actuate Corporation¹¹, al mismo tiempo que fue donado a la Eclipse Foundation cuando el proyecto fue aprobado se convirtió en un proyecto de alto nivel dentro de la comunidad Eclipse, el 6 de octubre del mismo año. **(Wikipedia, 2014)**

El proyecto BIRT también recibió el aporte de la comunidad Innovent Solutions¹², e IBM en el año 2007 la división Tivoli de IBM adoptó BIRT como infraestructura para su producto Tivoli Common Reporting (TCR) este reporte produce informes históricos en recursos y procesos IT gestionados por Tivoli. **(Wikipedia, 2014)**

¹¹ **Actuate Corporation:** Empresa con sede en San Mateo, California, es conocida por su creación del código abierto Eclipse BIRT puesto en marcha por la Fundación Eclipse en 2004. **Fuente:** Tomado de Wikipedia, Actuate Corporation [http://en.wikipedia.org/wiki/Actuate_Corporation]

¹² **Innovent Solutions:** Es una empresa de consultoría, dedicada a la mejora de las decisiones a través de la gestión eficaz de la información de decisiones. **Fuente:** Tomado de Innovent Solution [<http://www.innoventsolutions.com/company-overview.html>]

BIRT es una herramienta para aplicaciones Java/J2EE, los pasos que ofrece para llevar a cabo la creación de un reporte son muy sencillos y claros, también ofrece variadas opciones desde la vista de “Eclipse Palette”, como listas, gráficos, tablas, imágenes y varios elementos visuales más, que se podrán agregar a la ventana de diseño. Otras opciones que se pueden resaltar de BIRT son los filtros, resúmenes, clases, grupos y otros más, esta herramienta de reportes para Java con código abierto, es la utilidad que se necesita para las aplicaciones y reportes de la web, se puede modificar y configurar desde sus bases estéticas y de diseño, hasta el Javascript.

Los objetivos de Birt son:

- Cubrir un ancho rango de necesidades de creación de informes dentro de una aplicación típica, abarcando desde informes operacionales o de empresa hasta procesamiento multi-dimensional analítico en línea (OLAP). El proyecto recibe soporte desde una activa comunidad de usuarios en BIRT Exchange y desarrolladores de Eclipse.org.
- Permite generar informes (con o sin parámetros) en formato HTML, PDF, XLS, DOC o PPT ya que el motor de BIRT es una biblioteca.
- Ofrecer la realización de tablas dinámicas cruzadas, los datos presentados pueden proceder de diferentes bases y consultas; estos informes pueden ser complejos e incluir varias tablas, gráficos avanzados e imágenes.
- Integrar el motor de BIRT en cualquier aplicación desarrollada en Java, ya sea en una aplicación web o en una aplicación de tipo “cliente pesado”.
- Integrar los informes BIRT en la solución Pentaho a través de BIRT-Viewer.
- Disponer de una herramienta de diseño de informes que permita crear y modificar informes de forma gráfica, introduciendo elementos (tablas, gráficos, etc.) en un informe en blanco.
- Integrar en forma de plugin en la herramienta de desarrollo Eclipse, aunque también se puede utilizar como una aplicación autónoma.

- Permitir crear rápidamente informes complejos y controlar su formato.

BIRT (Business Intelligence and Reporting Tools que significa inteligencia de negocios y herramientas de reporte) tiene 2 componentes principales:

1. Un diseñador de informes visuales conocido como **Birt-Designer** dentro de Eclipse IDE para crear informes BIRT que es un componente de rutina para generar informes que pueden correr en cualquier entorno Java, esta interfaz tiene mucha similitud con Dreamweaver, aquí se hacen las respectivas conexiones a bases de datos, scripts, Servicios Web, etc., para que todo sea almacenado en un archivo XML con un extensión .rptdesing.
2. Existe también un visualizador conocido como **Birt-Viewer** este incluye un motor de gráficos que está integrado en el diseñador de informes, trabaja bajo un servidor de servlets o paginas jsp como por ejemplo Tomcat, el resultado es presentar varios formatos: html, pdf, cvs, Excel, para que puedan trabajar con cualquier otro lenguaje, por ejemplo php únicamente ingresando la URL que se genera localmente para que ésta haga el trabajo de crear el reporte.

3.1.2 ATTIVIO

Desde el año 2008 la empresa Attivio Inc. ha sido reconocida por Red Herring, IDC, Gartner, KMWorlds, como una de las empresas innovadoras y con mayor potencial para brindar valor a las organizaciones que han gastado su dinero en datawarehouse tradicionales y otras plataformas de vieja generación. Es una empresa privada de software empresarial que produce y vende información de acceso unificado, que permite a los usuarios encontrar todo tipo de información con una sola consulta. **(Wikipedia, Attivio, 2014)**

Attivio, Inc. lanza al mercado Active Intelligence Engine que es una marca comercial encargada de la gestión de inteligencia de negocios, resuelve el problema de grandes cantidades de información aislada, de tal manera que el usuario final mediante la integración de búsqueda empresarial, inteligencia de negocios y capacidades de automatización de procesos puedan dar respuestas y generar reportes para la toma de decisiones empresariales. **(Wikipedia, 2014)**

Ésta plataforma de acceso a la información unificada de Attivio, Active Intelligence Engine (AIE) define el impacto comercial de información de los usuarios finales, para que así ellos puedan desarrollar estrategias de negocios, resolver desafíos críticos del mercado actual y cumplir con su visión de ser competitivos frente a las grandes empresas de éxito.

La finalidad principal de Attivio es la de integrar almacenes dispares de datos estructurados, ofreciendo al usuario final capacidades de búsqueda intuitiva, Active Intelligence Engine integra las herramientas disponibles para inteligencia empresarial y para grandes cantidades de datos con el fin de mostrar información valiosa para la toma de decisiones, a través del método de acceso que mejor se adapte a las capacidades técnicas de cada usuario y a sus prioridades.

Esta plataforma se puede implementar de manera local o en la nube, los objetivos de esta plataforma son:

- Integrar la información estructurada o no estructurada, en cualquier formato, desde cualquier repositorio, dentro o fuera de la empresa.
- Interrelacionar la información a través de fuentes en tiempo de consulta real, permitiendo que el sistema sea ágil y mantenga relaciones de datos.
- Acceder a la información en cualquier nivel de encriptación o registro, mediante la herramienta de inteligencia de negocios preferida o a través de herramientas de visualización de datos mediante bases de datos de SQL.
- Analizar el contenido y los datos de todos los almacenes de información.

Active Intelligence Engine elimina los sobre abultamientos de información, esto permite a las empresas acceder rápidamente a nuevas necesidades de información, como resultado los usuarios obtienen nuevos reportes en tiempo real sobre la marcha de su negocio que ayudan a la buena toma de decisiones mejorar los ingresos, reducir costos y aumentar la competitividad.

3.1.2.1 Características de Active Intelligence Engine

Índice híbrido para almacenar cualquier tipo de información

- Active Intelligence Engine abarca todos los tipos de datos estructurados y no estructurados, la construcción de un esquema menos índice de información unificado que se puede acceder con una sola consulta, ayudando a obtener una visión más profunda y detectar patrones complejos.
- A diferencia de un almacén de datos tradicional, no requiere que las relaciones entre los datos estén definidas antes de la ingestión.
- Active Intelligence Engine también abarca datos estructurados en base a tablas, conservando la estructura lógica y la integridad de los datos a diferencia de otras soluciones de información como es eliminar la normalización de bases de datos, lo que disminuye en gran medida su valor de negocio.
- Es compatible con SQL.
- Actualizaciones en tiempo real.
- Reconocimiento óptico de caracteres.
- Servicio de búsqueda.

Consulta en tiempo real de toda la Información

Active Intelligence Engine permite la selección de las relaciones entre tablas, conectar los campos comunes para representar a uno-a-uno, uno-a-muchos y muchos-a-muchos, con el objetivo de mejorar la ejecución de los procesos en tiempo de consulta, recuperación de fuentes para garantizar que la información este actualizada en todo momento.

Compatibilidad con SQL

AIE admite SQL que es el lenguaje de consulta universalmente adoptado por las herramientas de análisis de datos; SQL permite precisión y complejidad más allá de las capacidades de las consultas en los motores de búsqueda y utiliza operadores estándar para aplicar las capacidades analíticas, a través de diferentes fuentes.

Actualizaciones en tiempo real

Active Intelligence Engine está diseñado para el funcionamiento en tiempo real; se pueden agregar nuevos documentos o registros para ser puestos a disposición de la búsqueda de lotes de 100 o 1.000 en menos de dos segundos, esta ventaja permite que el acceso a la información sea más rápida.

Integrales capacidades de búsqueda

AIE ofrece sugerencias de ortografía en función del contenido o diccionarios indexados, sinónimos y expansión siglas, etc., admite métodos de búsqueda estándar, entre los más comunes tenemos: palabra clave, simbología de sintaxis, búsqueda de frases, comodines, búsqueda difusa, etc.

Interfaz de usuario amigable

AIE proporciona una variedad de interfaces, incluyendo cuadros de mando, informes y resultados de búsqueda, en cualquier tipo de interfaz, incluyendo simples resultados de la búsqueda, como correos electrónicos, registros y otros archivos de texto que también se puede visualizar en respuesta a una sola consulta para obtener información sobre un tema.

Métodos de navegación para facilitar la búsqueda

AIE soporta la última navegación y herramientas de descubrimiento para simplificar la búsqueda de información, corrige la ortografía y sugiere cambios durante la entrada de consulta de los usuarios.

Modelo de seguridad activa

AIE acelera el permiso de actualización de almacenamiento de documentos, grupos, usuarios y listas de control de acceso en una tabla de seguridad por separado en el índice.

Configuraciones Certificadas Multi-Node

AIE admite configuraciones certificadas de las implementaciones de los clientes más comunes: alta disponibilidad para consultar, indexación garantizada, que permite la adición de nuevos nodos de procesamiento de contenido de cálculo intensivo, tales como extracción de texto y de OCR.

Fácil de instalar, configurar y utilizar

AIE es flexible, personalizable y escalable. Los administradores pueden integrar rápidamente las nuevas fuentes de información, utilizando interfaces web mientras está en marcha. Attivio abarca algunas herramientas asociadas a su marca como por ejemplo Cliksoft, QlikView.

Qlikview

Una de las soluciones que presenta Attivio es QlikView, que es una plataforma para la creación de aplicaciones analíticas desarrollada para la toma de decisiones, bien sea empresarial o institucional, desde cualquier fuente de datos, su objetivo principal es reducir los tiempos de desarrollo, formación e implantación de la solución, QlikView brinda a las empresas un sistema de business intelligence que puede crecer de acuerdo con la estrategia de negocio que tiene la empresa permitiendo la gestión diaria desde el primer momento de implementada la solución basada en cubos OLAP. **(EOS, 2014)**

Se describen los proyectos más destacados en los que se han desarrollado e implementado casos de Éxito:

- **Fybeca.**- El levantamiento de los requerimientos los realizan en base a entrevistas, hitos de revisión y entregables, siguiendo las pautas de la metodología RUP.

- **Corporación Maresa.**- El objetivo fue analizar y optimizar el proceso de alimentación de información del área de Marketing, para analizar las tendencias del mercado automotriz.
- **Corporación Maresa.**- Aquí se realizó la implantación del proyecto Radar2 (Geomarketing), con el objeto de identificar la densidad poblacional de los clientes, encontrar los centros de gravedad de la corporación y obtener la ubicación geográfica óptima de futuros servicios a nivel nacional.
- **Cidescol.**- Aquí se hizo un desarrollo de análisis de información sobre el sistema Mastersoft con Qlikview, ésta información está relacionada al análisis de clientes en el área de seguros con sus respectivas comparativas anuales de renovación.
- **Liderseg.**- Desarrolló el sistema de análisis de información sobre el sistema Mastersoft con Qlikview.
- **Industrias Omega.**- Primeramente se desarrolló con business intelligence una aplicación de ventas, control de vendedores y clientes, al momento de haber tenido toda esta información codificada y analizada prácticamente los reportes permitían acceder a la información de facturación, clientes y vendedores lo que les permitió realizar un análisis a nivel agregado y a detalle con gráficos generados en Dashboards, e indicadores de tablas.

3.1.3 PENTAHO

La Suite de Pentaho Business Intelligence es un paquete de software libre que nos permite generar inteligencia empresarial, este paquete incluye otras herramientas de software libre integradas para generar informes, minería de datos, ETL, etc. (**Wikipedia, Pentaho, 2014**)

Pentaho fue creado por 5 fundadores en el año 2004, tiene su sede en Orlando, Florida, EE.UU; el objetivo es ofrecer un paquete de código abierto de Inteligencia de Negocios, los productos llamados Pentaho Business Analytics proporciona integración de datos, servicios OLAP, informes, cuadros de mando, minería de datos y capacidades ETL. (**Wikipedia, 2014**)

La suite Pentaho consta de dos ofertas:

- **La empresarial.-** Se obtiene a través de una suscripción anual que incluye los servicios de apoyo adicionales, beneficios para el cliente, soporte en el paquete en el cual está trabajando y,
- **La edición de la comunidad.-** Conocida como Enterprise Edition incluye características adicionales que no se encuentran en la edición de la comunidad, con frecuencia se encuentran productos add-on¹³, por lo general en forma de plugins.

3.1.3.1 Descripción

Pentaho ofrece buenas soluciones de software que se componen principalmente de una infraestructura de herramientas para el análisis de información y generación de informes, los mismos que están integrados con un motor de workflow¹⁴ de procesos de negocio, ésta plataforma es capaz de ejecutar las reglas de negocio necesarias en forma de procesos y actividades, la suite de Pentaho se puede definirla como una plataforma “orientada a la solución” y “centrada en procesos” que permite incluir todos los componentes requeridos para implementar soluciones basados en procesos. En la estructura web Pentaho presenta una lista organizada de productos: Reporting, Analysis, Dahnboards y Data Mining.

3.1.3.2 Productos de Pentaho

- **Pentaho Analysis Services:** es un servidor OLAP escrito en Java. Este servidor es compatible con MDX (expresiones multidimensionales) y el lenguaje de consulta XML sirve para el Análisis de la interfaz olap4j.
- **Pentaho Reporting:** Es un motor de presentación, que permite generar informes programáticos sobre la base de un archivo de definición XML, gracias a esta herramienta se ha podido generar muchas soluciones, por

¹³ **Add-On:** Conocidos como extensiones, plugins, snap-ins, etc, son programas que sólo funcionan anexados a otro y que sirven para incrementar o complementar sus funcionalidades. **Fuente:** Tomado de Wikipedia, Add-on [<http://es.wikipedia.org/wiki/Add-on>]

¹⁴ **Workflow:** O flujo de trabajo es el estudio de los aspectos operacionales de una actividad de trabajo: cómo se estructuran las tareas, cómo se realizan, cuál es su orden correlativo, cómo se sincronizan, cómo fluye la información que soporta las tareas y cómo se le hace seguimiento al cumplimiento de las tareas. **Fuente:** Tomado de Wikipedia, Flujo de Trabajo [http://es.wikipedia.org/wiki/Flujo_de_trabajo]

ejemplo informes, diseñadores de interfaz gráfica de usuario y asistentes tipo wizard¹⁵ un ejemplo claro de esta herramienta es el Generador de informes para OpenOffice.org que es un paquete de software libre.

- **Pentaho Data Mining:** Es una suite de software libre que usa estrategias de aprendizaje automático de máquina y minería de datos. Ésta herramienta cuenta con series de clasificación, regresión, reglas de asociación y algoritmos de clustering¹⁶, su objetivo es el de apoyar las tareas de análisis predictivo, esta herramienta forma parte del proyecto Weka¹⁷.
- **Pentaho Dashboard:** Es una plataforma integrada que permite proporcionar información sobre los datos almacenados en la base de datos multidimensional, esta herramienta permite ver informes, gráficos interactivos y cubos creados con las herramientas Pentaho Report Designer.
- **Pentaho para Apache Hadoop:** Esta herramienta es un conector de bajo nivel, que permite facilitar el acceso a grandes volúmenes de información que es manejado por el proyecto Apache Hadoop.

3.2 Comparativa de Herramientas

Al realizar una comparativa de 3 herramientas de software libre como su nombre lo indica, se realizará un estudio comparativo de sus características, ventajas y desventajas, enlistar cada una de sus fortalezas más importantes.

Se ha podido analizar los pros y los contras de estas 3 herramientas de software libre, para dicho análisis se ha tomado en cuenta el fácil manejo de la herramienta, las tecnologías soportadas, base de datos compatibles, tiempo de implementación, sistemas operativos sobre los cuales corre la arquitectura, seguridad, licenciamiento, la integración con herramientas externas de gestión, el

¹⁵ **Wizard:** Asistente de ayuda. **Fuente:** Tomado de Diccionario Reverso [<http://diccionario.reverso.net/ingles-espanol/wizard>]

¹⁶ **Clustering:** Los algoritmos de clustering son usualmente empleados para mejorar el rendimiento y/o la disponibilidad de la información en una minería de datos. **Fuente:** Tomado de Wikipedia, Clúster [[http://es.wikipedia.org/wiki/Cl%C3%BAster_\(inform%C3%A1tica\)](http://es.wikipedia.org/wiki/Cl%C3%BAster_(inform%C3%A1tica))]

¹⁷ **Weka:** Es una plataforma de software para aprendizaje automático y minería de datos escrito en Java y desarrollado en la Universidad de Waikato, Weka es un software libre distribuido bajo licencia GNU-GPL. **Fuente:** Tomado de Wikipedia, Weka [[http://es.wikipedia.org/wiki/Weka_\(aprendizaje_autom%C3%A1tico\)](http://es.wikipedia.org/wiki/Weka_(aprendizaje_autom%C3%A1tico))]

soporte de las herramientas y las plataformas compatibles, de acuerdo a la investigación realizada de las 3 herramientas Ativio, Birt, Pentaho, se ha logrado determinar ventajas y desventajas de cada una de ellas.

3.2.1 Etapas que se deben seguir para la comparativa de herramientas

Antes de empezar a realizar el estudio comparativo de las herramientas de software libre predefinidas, es importante cumplir con algunos requerimientos, mismos que permiten identificar la herramienta que cumple con todas o la mayoría de etapas propuestas a continuación:

3.2.1.1 Descripción de las herramientas que nos permiten elaborar datamarts.- En esta etapa se hará una investigación a fondo de cada una de las herramientas con el objetivo de tener una idea clara y concisa en cuanto al uso, manejo, ventajas y desventajas; con el resultado obtenido se avanza con el siguiente paso.

3.2.1.2 Definición de características a evaluar de las 3 herramientas de software libre.- A continuación, se hará un resumen de las características más importantes que posee cada una de las herramientas a estudiarse para el desarrollo del datamart, esta etapa es muy importante ya que de la selección acertada de la herramienta depende el éxito o fracaso.

3.2.1.3 Evaluación de las herramientas de software libre.- Esta etapa permite describir de manera explícita las características de cada una de las herramientas a través de un cuadro comparativo, mediante la suma de los checks asignados a las ventajas y una X a las desventajas es como se va a determinar el total de cada una de las herramientas evaluadas.

3.2.1.4 Selección de la herramienta adecuada de software libre.- La herramienta seleccionada es la que tiene mayor sumatoria en checks de acuerdo a las características y ventajas determinadas en la etapa anterior.

3.2.2 Resultados

Luego de haber seguido y cumplido cada una de las distintas etapas descritas anteriormente para la obtención de resultados, a continuación se aplicarán los siguientes pasos para determinar el resultado definitivo:

Definición de las características a evaluar.- Es necesario explicar que para la realización del cuadro comparativo de herramientas de software libre se tomó en cuenta las características que se describen a continuación:

- **Plataforma:** La investigación está dirigida para pequeñas y medianas empresas sobre ese concepto se evaluó que las 3 herramientas sean multiplataforma o compatibles con plataformas Linux y Windows.
- **Definición de dimensiones:** La herramienta seleccionada deberá permitir que se definan dimensiones sobre la base de datos transaccional.
- **Datamining:** Luego de haberse realizado el proceso ETL, diseñado y publicado el cubo, debe permitir realizar minería de datos, que es muy importante al momento de implementar inteligencia de negocios.
- **Generación de dashboards¹⁸:** La herramienta debe integrarse perfectamente con la información para la generación de cuadros de mando, luego de haberse generado la minería de datos.
- **Servidor de aplicaciones:** El servidor de aplicaciones gestiona la comunicación entre el código fuente de la base de datos y los cubos OLAP, es importante que la herramienta use un servidor de aplicaciones y al mismo tiempo permita la generación de cubos OLAP.
- **Espacio de memoria y tiempo de respuesta en tiempo real:** Es importante que la herramienta optimice el consumo de recursos y su tiempo de respuesta ante cualquier consulta, sea en el menor tiempo posible.
- **Reportes:** Debe permitir la generación de reportes luego de generada la información.

¹⁸ **Dashboards:** Cuadros de mandos de control.

- **Base de datos compatibles:** La herramienta adecuada debe ser compatible con bases de datos tanto libres como con licenciamiento.
- **Seguridad:** La herramienta debe permitir la gestión y encriptación de la información de manera adecuada y segura.
- **Soporte.-** La herramienta deberá contar con soporte, actualizaciones de versiones para que se encuentre al día en el mercado.
- **Licenciamiento.-** La herramienta deberá permitir trabajar con versiones completas.

Si bien es muy importante tener en cuenta estos requerimientos para que la herramienta seleccionada sea la más adecuada al momento de realizar la implementación de inteligencia de negocios, pero también es importante que la selección cumpla con los requerimientos tanto del usuario como de la persona a desarrollar y ayude en el cumplimiento de las necesidades propuestas.

3.2.3 Cuadro comparativo de herramientas de software libre

	ATTIVIO (AIE)	BIRT	PENTAHO
SEGURIDAD	X	X	✓
LICENCIA OPEN SOURCE	✓	✓	✓
SSOO COMPATIBLES	Linux, Mac	Linux, Windows	Linux, Windows
ETL	X	X	✓
CUBOS OLAP	X	✓	✓
REPORTING	✓	✓	✓
VERSIONES COMMUNITY Y EMPRESARIAL	X	X	✓
ESPACIO EN MEMORIA	Ocupa mucho espacio	Ocupa bastante espacio	Ocupa poco espacio
WIZARDS E INTERFACES GRAFICAS	✓	X	✓
PERMITE DEFINIR DIMENSIONES	X	X	✓
DATAMINING	X	✓	✓
BASE DE DATOS COMPATIBLES	MySQL	MySQL	MySQL, POSTGRES, ORACLE, SQL
SERVIDOR DE APLICACIONES	X	X	✓
DASHBOARDS	✓	✓	✓
PLANIFICACIÓN DE TAREAS EN TIEMPO REAL	✓	X	✓
TOTAL	5	5	12

Tabla 3.1: Comparativa de Herramientas
Fuente: Elaboración propia

3.3 SELECCIÓN DE LA HERRAMIENTA ADECUADA

El proceso de selección de una herramienta adecuada para el desarrollo de business intelligence puede resultar complicado y tedioso, debido a que existen múltiples alternativas en el mercado actual, para lo cual se hace muy difícil poder argumentar razones sustentables frente a la decisión tomada.

Para tener una visión más amplia de cual herramienta elegir es necesario investigar y elaborar cuadros comparativos entre las mismas herramientas para así determinar ventajas y desventajas entre ellas, cada herramienta tiene características diferentes ya sea en plataforma, alcance, funcionalidad, tecnología, arquitectura, valor agregado, etc., lo que permite notar claramente sus fortalezas en algunos temas y ser deficientes en otros.

Hay que tomar en cuenta que muchas herramientas de inteligencia de negocios no son orientadas a los reportes operacionales, no todas las empresas están listas para implementar herramientas analíticas, muchas veces es conveniente iniciar por automatizar su reporte para facilitar su operación, madurar en esta etapa y posteriormente seguir con las etapas analíticas y de minería de datos.

Para el desarrollo del prototipo del presente trabajo de grado se escoge como herramientas para el desarrollo del prototipo a Pentaho para desarrollo de cubos y generación del datamart y a PostgreSQL para el diseño de la base de datos.

3.4 SELECCIÓN DEL PROCESO

Los profesionales de tecnologías de información y los gerentes de las pequeñas y medianas empresas tienen responsabilidades entre ellas la de tomar una decisión acertada en general si desarrollar una aplicación o comprar una herramienta, hay que tomar en cuenta muchos factores entre los más importantes están: los recursos económicos, disponibilidad de tiempo de parte del personal de desarrollo y el plazo para ejecutar el proyecto planteado, además hay que tomar en cuenta la velocidad a la que evoluciona el mercado de la tecnología.

Escoger una herramienta de business intelligence que satisfaga las necesidades de los usuarios en cuanto a funcionalidades, arquitectura y costo no es una tarea sencilla, para el personal de desarrollo seguir una metodología, e involucrar al personal adecuado es muy importante, ya que tiene que manejar dos grupos de usuarios los generadores de información y los consumidores de información. Es importante tomar en cuenta dos tipos de procesos para seleccionar la solución de inteligencia de negocios más adecuada:

3.4.1 Proceso informal: cuando la empresa no establece un proceso formal, en la mayoría de casos el resultado es que no se obtengan los resultados esperados.

3.4.2 Proceso formal: es un proceso que se lo ejecuta paso a paso de acuerdo al planteamiento de los objetivos fijados en la empresa se inicia con el planteamiento del proyecto, definición de requerimientos hasta llegar a la conclusión del mismo con grandes resultados.

Es necesario tomar en cuenta ciertos criterios a la hora de evaluar y seleccionar una herramienta de software como las funcionalidades de las consultas, informes, administración, monitoreo, fácil colaboración de los proveedores de información, etc., la herramienta debe soportar nuevas versiones de software; sin embargo la evaluación de la mejor decisión debe realizarse bajo un esquema que involucre a cada uno de los stakeholders¹⁹, este proceso debe tomarse en cuenta a cada instante ya que el impacto financiero y operativo va ligado directamente a la estrategia corporativa de la empresa.

3.5 GRANULARIDAD DE LA INFORMACIÓN

Uno de los procesos que más complejidad tienen dentro del área de desarrollo de software es la selección de componentes que integran el proyecto, este es un tema que involucra a desarrolladores y a usuarios finales, en algunos casos estos componentes pueden ser demasiado grandes, complicando el proceso de selección y adquisición del proyecto.

¹⁹ **Stakeholders:** Son las personas interesadas o involucradas en un proyecto determinado, y que necesitan generar una solución. **Fuente:** Tomado de Wikipedia, Stakeholders, [<http://es.wikipedia.org/wiki/Stakeholder>]

Definición de Granularidad de Información

La granularidad es el nivel de detalle de una base de datos dimensional en un datawarehouse, para el modelamiento es muy importante determinar el nivel de granularidad de la información, esto afecta directamente en el tamaño de la base de datos a analizar. **(Wikipedia, 2014)**

3.5.1 Determinación de la granularidad de la tabla de hechos

Para determinar la granularidad de una tabla de hechos es necesario analizar los requisitos del proceso de negocio, dependiendo de la naturaleza del negocio los documentos con frecuencia incluyen información que se puede utilizar para definir la granularidad, esta información ayuda a identificar las dimensiones y medidas de los modelos dimensionales.

El nivel de granularidad que se elija es la base de cada modelo dimensional, éste determinará el nivel de detalle de la información que puede haber disponible en el modelo dimensional necesaria para cada área del negocio. Es importante saber qué nivel de detalle está asociado con un cliente, un producto o un proveedor.

CAPÍTULO IV

DISEÑO PROTOTIPO

4.1 IDENTIFICACIÓN DE LAS DIMENSIONES PARA EL DISEÑO DEL PROTOTIPO

Es importante determinar la granularidad de la tabla de hechos para poder identificar las dimensiones definitivas, primero hay que identificar y elaborar las tablas de dimensiones para el éxito de la tabla de hechos o FACT, a continuación se describirá los aspectos a tomar en cuenta para el diseño de las dimensiones:

- Nombre que defina e identifique claramente la dimensión
- Conexión con la base de datos transaccional
- Calidad y precisión de los datos
- Definición de claves primarias
- Información de IP, nombre ETL, fecha del sistema, para hacer un registro histórico de posibles errores en una tabla de auditoria
- Conexión con la base de datos dimensional
- Identificación de dimensiones a elaborar de acuerdo al área de negocio

Las tablas de dimensiones están conformadas de columnas que contienen los registros en la tabla de hechos, estas tablas dimensionales son más pequeñas, son tablas de búsqueda sin normalizar a las que se hace referencia al definir las consultas en la base de datos transaccional. Para el caso del prototipo que es un cubo académico se necesitarán las siguientes dimensiones:

- Dimensión alumno
- Dimensión profesor

- Dimensión materia_curso
- Dimensión curso_paralelo
- Dimensión carrera_materia
- Dimensión tiempo

4.2 IDENTIFICACIÓN DE LOS HECHOS O MÉTRICAS

Los hechos o métricas son las medidas que generalmente están representadas por cantidades, es decir son aquellas medidas cuantificables que miden los procesos del negocio que están encaminados al área del negocio. Los hechos son analizados por las tablas dimensionales que componen el modelado dimensional. Hay que distinguir que los hechos contienen columnas de datos llamadas métricas y las dimensiones están compuestas por columnas que se denominan niveles de jerarquías. Por ejemplo se cita la jerarquía **tiempo** que contiene mes, bimestre, semestre, año y como una métrica sería la **nota** en este caso las notas que corresponden a primer, segundo bimestre y supletorio para luego de allí sacar un promedio correspondiente a un semestre y tener información valiosa para analizar.

4.3 IMPLEMENTACIÓN

Para la implementación del datamart es necesario tomar en cuenta la forma como se estructura el almacenamiento de los datos, a partir de esto se debe tomar en cuenta las pautas para implementar una arquitectura de datawarehouse, la arquitectura de software que se adopte se debe de implementar de la manera que satisfaga las necesidades del usuario siendo un factor muy importante en la efectividad del datamart.

En esta parte de la implementación se detallará cómo se construyó las dimensiones, el modelo de datos, el mapeo de la información y como se estructurará el cubo académico para el aplicativo, también se detalla las tablas dimensionales que conforman la base de datos multidimensional que fueron creadas para la extracción y almacenamiento de información identificando sus respectivas áreas de negocio.

4.3.1 Diseño del prototipo

Para el diseño del prototipo se tomará en cuenta tanto el modelado físico como el modelado lógico, es muy importante tomar en cuenta ambos tipos de modelamiento ya que están conformados de herramientas a utilizar, el modelo de la base de datos, la base de datos dimensional, etc., la base de datos creada para ser utilizada por el datamart se denominó: TESISDW,

4.3.2 Herramientas de Software Libre que se utilizarán

Las herramientas de software libre que se utilizarán en el diseño del prototipo se detallan a continuación:

Software	Descripción
Pentaho BI server	Es un servidor OLAP.
Pentaho Report Designer	Es una herramienta para sacar reportes.
Pentaho Data Integration	Herramienta que nos ayuda en la integración de datos, nos permite realizar proceso ETL.
Pentaho Schema Work Bench	Herramienta que nos sirve para diseñar cubos dimensionales.
Base de datos PostgreSQL	Repositorio del datawarehouse.
DBDesigner 4	Herramienta que ayuda en el diseño de la base de datos.

Tabla 4.1: Herramientas de software libre para desarrollar un datawarehouse
Fuente: Elaboración propia

4.3.2.1 Pentaho Open Source Business Intelligence

Ésta herramienta Open Source es una alternativa recomendada para inteligencia de negocios.



Figura 4.1 Logotipo de la Suite Pentaho
Fuente: [<http://espritsur.com.ec/category/top-posts/>]

Ésta suite proporciona muchas funcionalidades entre ellas están: análisis OLAP, cuadros de mando, minería de datos, reporting, integración de datos y plataforma BI, esta herramienta se ha convertido en la más utilizada por las pequeñas y medianas empresas, por ser de código libre. La plataforma permite realizar las siguientes actividades:

- **Analizar:** Explorar la información de manera interactiva y de forma ágil.
- **Informar:** Acceder a la información y proporcionar información a toda la empresa.
- **Integrar:** Integración de la información desde el lugar donde se encuentre y desde varias fuentes.
- **Investigar:** Permite realizar minería de datos para descubrir esquemas ocultos y definir indicadores de tendencia a futuro.
- **Sintetizar:** Acceder de manera inmediata a través de cuadros de mando mostrando de manera visible medidas y frecuencias.

4.3.2.2 Pentaho Report Designer

Esta herramienta es una recopilación de software de código abierto que se enfoca principalmente en la creación, generación y distribución de reportes de manera organizada de todas las fuentes de datos.



Figura 4.2 Logotipo de la Suite Pentaho Reporting
Fuente: [<http://gravitar.biz/pentaho/caracteristicas-pentaho/>]

Este módulo está basado en el proyecto JFreeReport, permite distribuir los resultados del análisis de información en varios formatos como pdf, txt, html, etc., e inclusive tiene la opción para imprimir, posee un diseñador gráfico y plantillas de reportes que permiten mejorar la presentación de la información en los reportes.

4.3.2.3 Pentaho Data Integration

Es una muy buena herramienta gráfica que permite realizar procesos de ETL con la información de la base de datos transaccional para luego obtener tablas dimensionales.



Figura 4.3 Logotipo de la Suite Pentaho Data Integration

Fuente: [<https://www.sivsa.com/site/es/productos-y-servicios/soluciones/business-intelligence/item/23?start=4>]

Permite realizar migraciones de información entre bases de datos, exportación de información desde una base de datos a un archivo plano, cargar información en grandes volúmenes hacia cualquier base de datos y una muy buena integración de las aplicaciones. Kettle está conformado de los siguientes paquetes:

- **Spoon**

Es el diseñador gráfico que permite realizar el proceso de las transformaciones ETL. Esta interfaz gráfica permite al usuario diseñar las dimensiones de información desde la base de datos transaccional a la base de datos dimensional a través de iconos fáciles de configurar.

- **Pan**

Es un motor interno que ayuda en la transformación de la información, realiza muchas tareas como leer, escribir y organizar la información desde varias fuentes de datos.

- **Kitchen**

Este programa permite ejecutar las transformaciones hechas en la herramienta gráfica Spoon, su funcionamiento es sencillo y realiza una secuencia de pasos para completar la transformación y carga de información.

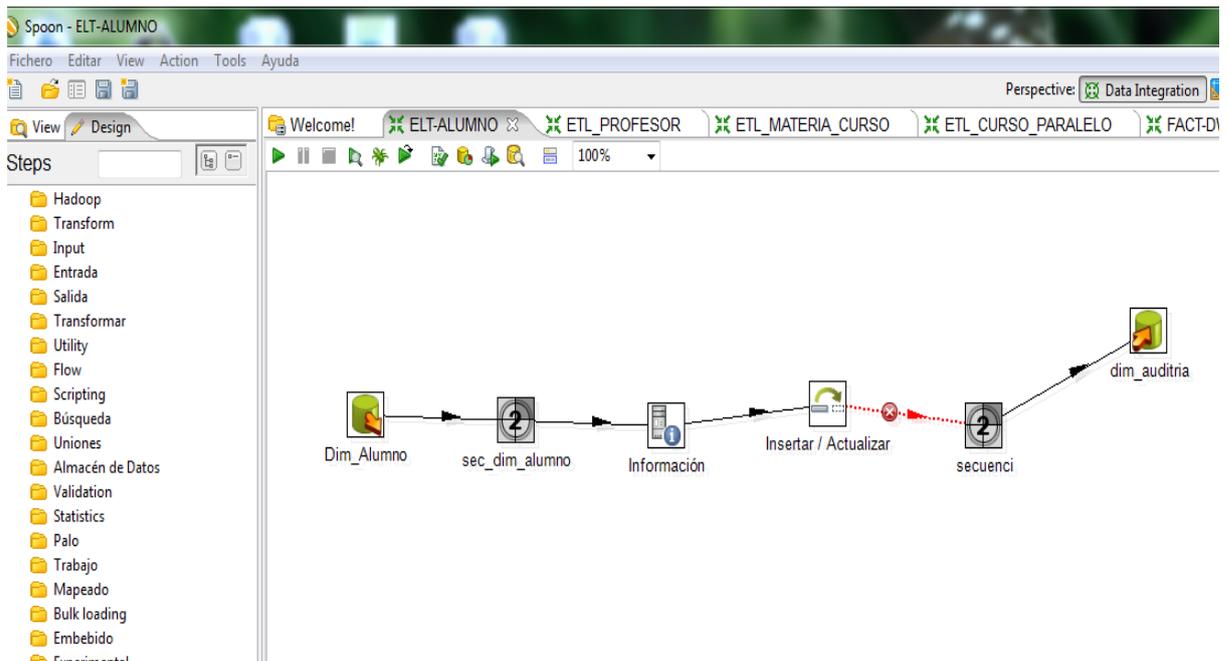


Figura 4.4 Herramienta Spoon para generar ETL
Fuente: Elaboración propia

4.3.2.4 Servidor Mondrian

El servidor OLAP open source Mondrian contiene como núcleo un archivo Jar, este archivo trabaja como un conector JDBC que ayuda hacer conexiones con las distintas bases de datos y permite ejecutar consultas SQL de la base de datos transaccional, facilita la gestión de comunicación desde una aplicación OLAP y una base de datos que contiene datos fuente, Mondrian está empaquetado de dos formas: como un paquete WAR que está conformado de Jpivot y el otro paquete conformado de Jpivot con una base de datos Derby. **(Hide Julian, 2014)**

Mondrian es compatible con las siguientes bases de datos: apache Derby, Firebird, Hsqldb, IBM DB2, Access, SQL Server, MySQL, Oracle, PostgreSQL, Sybase, una de las ventajas de este servidor es que reconoce de forma genérica el driver JDBC de cualquier base de datos inclusive si no aparece en la lista de compatibles y permite generar sentencias SQL de manera correcta.

4.3.2.5 Jpivot

Es un proyecto nuevo basado en JSP trabaja conjuntamente con Mondrian ya que combinados ambos se pueden realizar consultas MDX²⁰ y consultas desde elementos gráficos renderizados en un navegador web, la Suite Pentaho se ha adaptado a su patrón de diseño para poder diferenciar la interfaz original, entre sus características más importantes destacan:

- Capacidad de análisis interactivo por medio de un acceso web basado en Excel brindando una mayor funcionalidad, además tiene la posibilidad de ampliarse mediante el desarrollo.
- Está diseñada en estándares JDBC, JNDI, SQL, XML, MDX.

4.3.2.6 Pentaho Schema WorkBench

Es una interfaz gráfica de la suite Pentaho Analysis Services que permite elaborar, modificar y probar esquemas de cubos OLAP de manera gráfica, Mondrian procesa las solicitudes de MDX con los esquemas ROLAP estos esquemas pueden ser publicados en el servidor de business intelligence. Esta herramienta genera archivos de esquema XML, permite diseñar cubos de información con las respectivas dimensiones y medidas del cubo dimensional, generado a través de la información de la base de datos transaccional; esta herramienta está orientada al programador que conoce la estructura del esquema de Mondrian, está conformada de dos áreas la primera es la que muestra una estructura jerárquica del esquema OLAP y la segunda área es la que muestra la edición de propiedades de cada uno de sus elementos. **(Wood Sherman, 2014)**

Proporciona las siguientes funcionalidades:

- Editor de esquema integrado con el origen de datos subyacente para su validación.
- Consultas MDX de prueba en contra de esquema y la base de datos de pantalla.
- Revisar bases de datos subyacente estructura de pantalla.

²⁰ MDX: Es el lenguaje con el que trabaja el servidor OLAP Mondrian.

4.3.2.7 Base de datos PostgreSQL

PostgreSQL es una base de datos de código abierto dirigida por una comunidad de desarrolladores, esta base de datos solo posee un Storage Manager a diferencia de MySQL que tiene 5, está compuesta de algunos módulos que ayudan en la administración de las transacciones y el acceso a los objetos de la base de datos. Se rige bajo estos 3 lineamientos: manejo de transacciones, mantenimiento de una versión histórica de los datos y el aprovechamiento de las características que el hardware ofrece.

Entre algunas características importantes tenemos:

- Permite realizar copias de seguridad en caliente (online)
- Integra en su estructura herencia entre tablas, es por ello que se la podría considerar entre los gestores objeto – relacionales.
- Es robusta, soporta varios tipos de datos (tipo fecha, moneda, cadenas de bits, gráficos, IP, etc.)
- Es una base de datos Unicode.
- Su licencia es BSD (licencia de software libre permisiva).
- Compatible en los sistemas operativos Linux, Unix, Windows.
- Regionalización de cada columna.
- Control concurrente multi-versión.
- Posee algunos métodos de autenticación.
- Facilidad en la sintaxis del lenguaje SQL.
- Ideal para tecnologías web.
- Fácil administración en usuarios y privilegios.
- Escalable y confiable.

4.3.2.8 DBDesigner 4

Es una herramienta que corre bajo las plataformas Windows y Linux KDE/Gnome, permite el diseño de base de datos, modelamiento, creación y mantenimiento en el mismo entorno, tiene una interfaz amigable al usuario, es de código abierto y está publicada bajo la licencia GNU GPL, permitiendo al usuario descargarse el código fuente y realizar modificaciones, fue desarrollado por el programador Michael Zinner para la plataforma FabForce.net que usa Delphi y Kylix para ayudar a los usuarios de MySQL. **(fabFORCE.net, 2014)**

Una de las ventajas de esta herramienta es que cuenta con un motor de ingeniería inversa para base de datos MySQL, Oracle, MSSQL, también cuenta con dos interfaces conmutables: el modo diseño y el modo de consulta, los tipos de datos los define el usuario al momento del modelamiento, permite que el proyecto se guarde en formato XML para mantener visible la información.

4.3.3 Modelamiento de datos lógico

Para el modelamiento lógico del datamart se deberá tomar en cuenta cómo están diseñadas las dimensiones, que campos contiene cada dimensión, cómo está estructurada la tabla de hechos, parámetros técnicos a utilizar para identificar a cada uno de los hechos y medidas.

En este tipo de modelamiento se debe de tomar en cuenta la estructura que tiene cada tabla dimensional con el objetivo de construir un esquema de información basándose en un modelo de base de datos específico, independiente y concreto, cumplir con etapas y tareas de cada uno de los procedimientos a realizar. Luego de terminar el modelamiento lógico se da paso al modelamiento físico o dimensional.

4.3.3.1 Definición de parámetros

Para mayor facilidad se definió parámetros para identificar los nombres de las dimensiones, tablas de hechos, secuencias, herramientas de búsqueda, herramientas de información, etc., con la finalidad de ayudar al desarrollador a entender de mejor manera cada una de las herramientas. A continuación se definirán y asignarán los siguientes términos:

- **DIM.**- Término utilizado para definir las dimensiones.
- **FACT.**- Término utilizado para definir a las tablas de hechos.
- **AUD.**- Término utilizado para definir a las tablas de auditoría.
- **SEC.**- Término utilizado para definir a las secuencias.
- **Área de negocio.**- Término utilizado para definir a una entidad.
- **SK.**- Término utilizado para definir a una clave subrogada de tipo Num, esta puede ser generada por el secuencial de la base de datos o por alguna herramienta ETL.
- **Medida.**- Término utilizado para definir a las operaciones numéricas que se realiza en un campo.

4.3.3.2 Definición de parámetros para las tablas dimensionales

Para la creación del Datamart se diseñarán tablas con parámetros definidos para una acertada identificación y elaboración de las dimensiones que se crearán. Se seguirá los parámetros anteriormente descritos para definir a cada una de las tablas:

DEFINICIÓN	ABREV.	VISTA	DESCRIPCIÓN
Para las tablas de dimensiones	DIM_	DIM_ALUMNO	DIM_ÁREA DE NEGOCIO, máximo 25 caracteres
Para las tablas de hechos	FACT	FACT_DATOS_FAC	DIM_ÁREA DE NEGOCIO, máximo 40 caracteres
Para las tablas de auditoría	AUD_	AUD_ÁREA DEL NEGOCIO	AUD_ÁREA DE NEGOCIO, máximo 15 caracteres
Para las secuencias	SEC_	SEC_ÁREA DEL NEGOCIO	SEC_ÁREA DE NEGOCIO, máximo 25 caracteres

Tabla 4.2 Definición de parámetros para tablas y secuencias
Fuente: Elaboración propia

4.3.3.3 Definición de parámetros para los campos de las dimensiones

Es importante aclarar que las claves primarias de las dimensiones tengan una clave subrogada conocida como SK, esto se lo realiza mediante una secuencia en la base de datos. A continuación se definirá los parámetros para los diferentes campos en la siguiente tabla:

DEFINICIÓN	ABREV.	VISTA	DESCRIPCIÓN
Para las dimensiones las claves primarias son los SK	Alm_	ALM_SK	ÁREA DE NEGOCIO_SK, máximo 15 caracteres
Para los campos	Alm_	ALM_ALUMNO	Alm_Identificador, máximo 15 caracteres

Tabla 4.3 Definición de parámetros para campos de dimensiones
Fuente: Elaboración propia

4.3.3.4 Definición de parámetros para las secuencias

Las secuencias ayudan a generar las claves subrogadas SK para cada una de las dimensiones y la tabla de auditoría, estas claves subrogadas serán las claves primarias para cada dimensión, a continuación se describirán las secuencias creadas en el prototipo en la siguiente tabla:

DEFINICIÓN	ABREV.	VISTA	DESCRIPCIÓN
Secuencia para la dimensión alumno	SEC_	SEC_DIM_ALUMNO	SEC_DIM_ÁREA DE NEGOCIO
Secuencia para la dimensión profesor	SEC_	SEC_DIM_PROFESOR	SEC_DIM_ÁREA DE NEGOCIO
Secuencia para la dimensión tiempo	SEC_	SEC_DIM_TIEMPO	SEC_DIM_ÁREA DE NEGOCIO
Secuencia para la dimensión materia_curso	SEC_	SEC_DIM_MATERIA_CURSO	SEC_DIM_ÁREA DE NEGOCIO
Secuencia para la dimensión curso_paralelo	SEC_	SEC_DIM_CURSO_PARALELO	SEC_DIM_ÁREA DE NEGOCIO

Tabla 4.4 Definición de parámetros para secuencias
Fuente: Elaboración propia

4.3.3.5 Definición de parámetros para el proceso ETL

Para realizar el proceso ETL completo de una dimensión se utilizarán las siguientes herramientas:

- **Table input.-** Tabla de entrada de datos desde la base de datos de origen, se identificará de acuerdo al nombre de la tabla de origen.
- **Add sequence.-** Las secuencias llevarán el nombre que fue asignado anteriormente en la base de datos de destino.
- **Get System Info.-** Obtiene información del sistema.
- **Insert / Update.-** Procesará los datos para ser cargados en la base de datos destino.
- **Table output.-** Tabla de salida es la que contiene los datos de la tabla de auditoría.

HERRAMIENTA	NOMBRE ASIGNADO PARA ETL	DESCRIPCIÓN
Table input	DIM_ALUMNO	Esta tabla recibe los datos de la base de datos de origen extraídos mediante sentencias SQL.
Add sequence	SEC_DIM_ALUMNO SEC_AUD_	Secuencia para unir las dimensiones y la tabla de auditoría.
Get System Info	INFO_DIM_CLIENTE	Permite obtener la información del sistema para informes de auditoría.
Insert/Update	INSERT_DIM_ALUMNO	Permite el ingreso y la actualización de la información en la base de datos destino.
Table output	DIM_AUDITORÍA	Aquí se almacenan los datos para el informe de auditoría.

Tabla 4.5 Definición de parámetros para el proceso ETL
Fuente: Elaboración propia

4.3.3.6 Modelamiento de datos físico y representación de información

- **Modelado dimensional**

El modelo dimensional se dio lugar a través de la fase del diseño lógico, este modelo fue creado como propuesta para el diseño de datawarehouse con el objeto de que los usuarios que tienen datos empresariales hagan frente a los problemas de análisis multidimensional, cabe recalcar que en este modelo los datos están representados como si estuvieran en un espacio n-dimensional o cubo de datos, logrando de esta manera un análisis profundo de los hechos y las dimensiones creadas desde diferentes puntos de vista. El modelamiento dimensional surge de una adaptación del modelo relacional, así la información podrá ser almacenada en grandes bases de datos, de manera que los datos puedan ser extraídos y analizados mediante consultas OLAP, el modelo dimensional consiste en una sola tabla grande de datos que son descritos usando dimensiones y medidas. Los objetivos del modelamiento dimensional son:

- ✓ Presentar la información al usuario final en la forma más simple.
- ✓ Proveer información importante que se guarde como respaldo de los diferentes procesos generados.
- ✓ Brindar resultados rápidos y oportunos a los usuarios finales debido a las tablas sin normalizar que están involucradas con la creación de las dimensiones.
- ✓ Contar con una tabla de hechos normalizada y con tablas dimensionales sin normalizar.
- ✓ Ser adaptables a cambios.

- **Diagrama de origen de datos**

A continuación se describirá mediante un diagrama relacional las dimensiones y la tabla de hechos que contiene el proyecto con la base de datos de origen, para entender de mejor manera el origen de los datos que se utilizarán para la elaboración del datamart.

Para la presente investigación se utilizó el esquema estrella para el diseño, logrando de esta manera simplicidad en la implementación y velocidad en el acceso a la información y a las distintas jerarquías.

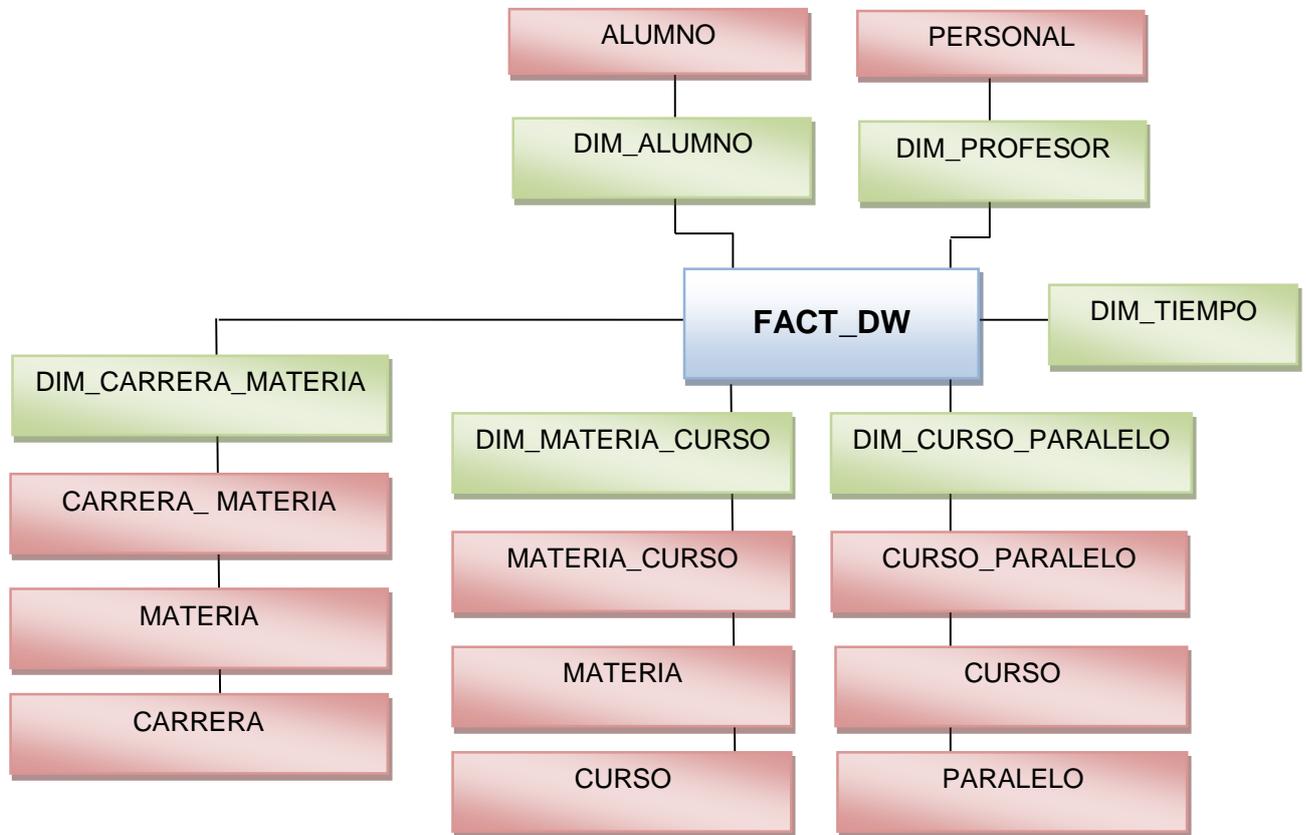


Figura 4.5 Diagrama origen de datos
Fuente: Elaboración propia

Las dimensiones se describirán a continuación:

Dimensión Alumno.- En ésta dimensión se cargan los datos básicos del alumno como nombres, cedula de identidad, género, etc., dimensión utilizada para analizar el indicador de alumnos.



Figura 4.6 Dimensión DIM_ALUMNO
Fuente: Elaboración propia

Dimensión Profesor.- En la dimensión profesor se carga la información básica como por ejemplo nombres, apellidos, género, etc., utilizada para analizar el indicador de docentes.



Figura 4.7 Dimensión DIM_PROFESOR
Fuente: Elaboración propia

Dimensión Materia_Curso.- En la dimensión Materia_Curso se carga la información de las materias que pertenecen a cierto curso.



Figura 4.8 Dimensión DIM_MATERIA_CURSO
Fuente: Elaboración propia

Dimensión Curso_Paralelo.- En la dimensión Curso_Paralelo se carga la información de los paralelos que pertenecen a cierto curso.



Figura 4.9 Dimensión DIM_CURSO_PARALELO
Fuente: Elaboración propia

Dimensión Carrera_Materia.- En la dimensión Carrera_Materia se carga la información de las materias que pertenecen a cada carrera.



Figura 4.10 Dimensión DIM_CARRERA_MATERIA
Fuente: Elaboración propia

Dimensión tiempo.- Es la que determina a que año, mes, semestre, etc., pertenece la información aquí se realizarán las consultas en función del tiempo.

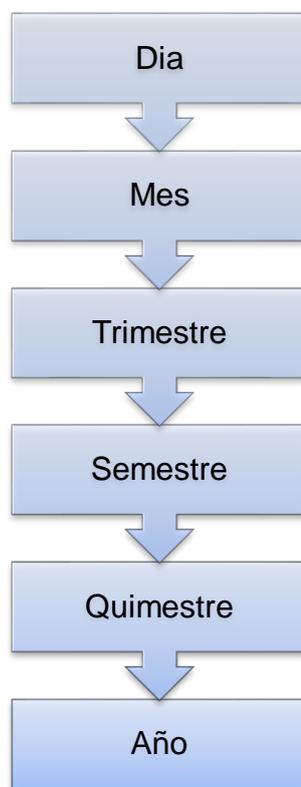


Figura 4.11 Dimensión Tiempo DIM_TIEMPO
Fuente: Elaboración propia

- **Mapeo de datos de origen**

Es importante realizar el mapeo de los datos de origen para que la persona en cargada de administrar el datawarehouse y sus respectivos datamarts tenga conocimiento de su estructuración, es de gran ayuda si llegaran a presentarse algunas fallas, para así evitar costos de mantenimiento y soporte.

En la siguiente tabla se describirá a continuación las tablas de origen de datos y su direccionamiento hacia las tablas dimensionales y la tabla de hechos que conforman el datamart.

TABLAS DE BASE DE DATOS DE ORIGEN	DESCRIPCIÓN	NOMBRE DE DIMENSIONES (DIM)	TABLA DE HECHOS (FACT)
ALUMNO	Esta tabla contiene información detallada de los alumnos.	DIM_ALUMNO	FACT_SISTEMAS_DW
PERSONAL	Esta tabla contiene información detallada de los profesores.	DIM_PROFESOR	FACT_SISTEMAS_DW
MATERIA_CURSO	Esta tabla contiene información detallada de las materias pertenecientes a cada curso.	DIM_MATERIA_CURSO	FACT_SISTEMAS_DW FACT_DATOS_DW
CURSO_PARALELO	Esta tabla contiene información detallada de los paralelos que pertenecen a cada curso.	DIM_CURSO_PARALELO	FACT_SISTEMAS_DW FACT_DATOS_DW
CARRERA	Esta tabla contiene información detallada de las carreras que pertenecen a la facultad.	DIM_CARRERA_MATERIA	FACT_DATOS_DW

Tabla 4.6 Mapeo de datos de origen
Fuente: Elaboración propia

- **Mapeo y detalle de la información de las dimensiones y hechos**

A continuación se describirá de manera detallada el origen y el destino de la información que conforman las dimensiones y la tabla de hechos que pertenecen al cubo a diseñar.

Nombre de dimensión: DIM_ALUMNO

Tipo de tabla destino: dimensional

Esquema: Public

Descripción: Dimensión que conforma el cubo ALUMNO_MATERIA, CARRERA_MATERIA

Origen			Destino				
Campo	Tipo de dato	Tam	Nombre Columna	Descripción	Tabla	FK	Esquema
	bigserial		alm_sk	Clave subrogada de la dimensión alumno	dim_alumno	alm_sk	public
id_alumno	bigserial		alm_id_alumno	Código del alumno	dim_alumno		public
nombres	character varying	200	alm_nombres	Campo donde se almacenan los nombres del alumno	dim_alumno		public
apellidos	character varying	200	alm_apellidos	Campo donde se almacenan los apellidos del alumno	dim_alumno		public
género	character varying	100	alm_genero	Campo donde se almacena el género del alumno	dim_alumno		public
cedula	character varying		alm_cedula_identidad	Campo donde se almacena la cédula del alumno	dim_alumno		public
FECHA_ETL	date		aud_fecha	Campo donde registra la fecha de carga del ETL	dim_auditoria		public
NOMBRE_ETL	character varying		aud_nombreetl	Campo donde registra el nombre del proceso ETL	dim_auditoria		public
IP	character varying		aud_ip	Campo donde registra la IP de la maquina donde se creó el ETL	dim_auditoria		public

Tabla 4.7 Mapeo de información de la dimensión DIM_ALUMNO

Fuente: Elaboración propia

Nombre de dimensión: DIM_PROFESOR

Tipo de tabla destino: dimensional

Esquema: Public

Descripción: Dimensión que conforma el cubo ALUMNO_MATERIA

Origen			Destino				
Campo	Tipo de dato	Tam	Nombre Columna	Descripción	Tabla	FK	Esquema
	bigserial		prof_sk	Clave subrogada de la dimensión profesor	dim_profesor	prof_sk	public
id_personal	bigserial		prof_id_profesor	Código del profesor	dim_profesor		public
nombres	character varying	200	prof_nombres	Campo donde se almacenan los nombres del profesor	dim_profesor		public
apellidos	character varying	200	prof_apellidos	Campo donde se almacenan los apellidos del profesor	dim_profesor		public
género	character varying	100	prof_genero	Campo donde se almacena el género del profesor	dim_profesor		public
cedula	character varying		prof_cedula_identidad	Campo donde se almacena la cédula del profesor	dim_profesor		public
FECHA_ETL	date		aud_fecha	Campo donde registra la fecha de carga del ETL	dim_auditoria		public
NOMBRE_ETL	character varying		aud_nombreetl	Campo donde registra el nombre del proceso ETL	dim_auditoria		public
IP	character varying		aud_ip	Campo donde registra la IP de la maquina donde se creó el ETL	dim_auditoria		public

Tabla 4.8 Mapeo de información de la dimensión DIM_PROFESOR

Fuente: Elaboración propia

Nombre de dimensión: DIM_MATERIA_CURSO

Tipo de tabla destino: dimensional

Esquema: Public

Descripción: Dimensión que conforma el cubo ALUMNO_MATERIA, CARRERA_MATERIA

Origen			Destino				
Campo	Tipo de dato	Tam	Nombre Columna	Descripción	Tabla	FK	Esquema
	bigserial		mat_sk	Clave subrogada de la dimensión materia_curso	dim_materia_curso	mat_sk	public
id_materiacursp	bigserial		mat_id_materiacurso	Campo donde se almacena el código de materia_curso	dim_materia_curso		public
id_curso	bigserial		mat_id_curso	Campo donde se almacena el código de curso	dim_materia_curso		public
id_materia	bigserial		mat_id_materia	Campo donde se almacena el código de materia	dim_materia_curso		public
num_credits	integer		mat_num_credits	Campo donde se almacena el número de créditos de la materia	dim_materia_curso		public
nombre	character varying	300	mat_nombre	Campo donde se almacena el nombre de la materia	dim_materia_curso		public
FECHA_ETL	date		aud_fecha	Campo donde registra la fecha de carga del ETL	dim_auditoria		sistema
NOMBRE_ETL	character varying		aud_nombreetl	Campo donde registra el nombre del proceso ETL	dim_auditoria		sistema
IP	character varying		aud_ip	Campo donde registra la IP de la maquina donde se creó el ETL	dim_auditoria		sistema

Tabla 4.9 Mapeo de información de la dimensión DIM_MATERIA_CURSO

Fuente: Elaboración propia

Nombre de dimensión: DIM_CURSO_PARALELO

Tipo de tabla destino: dimensional

Esquema: Public

Descripción: Dimensión que conforma el cubo ALUMNO_MATERIA, CARRERA_MATERIA

Origen			Destino				
Campo	Tipo de dato	Tam	Nombre Columna	Descripción	Tabla	FK	Esquema
	bigserial		cur_sk	Clave subrogada de la dimensión curso_paralelo	dim_curso_paralelo	cur_sk	public
id_curso	bigserial		cur_id_curso	Campo donde se almacena el código de curso	dim_curso_paralelo		public
curso	character varying	300	cur_curso	Campo donde se almacena el nombre del curso	dim_curso_paralelo		public
id_paralelo	bigserial		cur_id_paralelo	Campo donde se almacena el código del paralelo	dim_curso_paralelo		public
id_cursoparalelo	bigserial		cur_id_cursoparalelo	Campo donde se almacena el código de curso_paralelo	dim_curso_paralelo		public
nombre	character varying	300	cur_nombre	Campo donde se almacena el nombre del paralelo	dim_curso_paralelo		public
FECHA_ETL	date		aud_fecha	Campo donde registra la fecha de carga del ETL	dim_auditoria		public
NOMBRE_ETL	character varying		aud_nombreetl	Campo donde registra el nombre del proceso ETL	dim_auditoria		public
IP	character varying		aud_ip	Campo donde registra la IP de la maquina donde se creó el ETL	dim_auditoria		public

Tabla 4.10 Mapeo de información de la dimensión DIM_CURSO_PARALELO

Fuente: Elaboración propia

Nombre de dimensión: DIM_CARRERA_MATERIA

Tipo de tabla destino: dimensional

Esquema: Public

Descripción: Dimensión que conforma el cubo CARRERA_MATERIA

Origen			Destino				
Campo	Tipo de dato	Tam	Nombre Columna	Descripción	Tabla	FK	Esquema
	bigserial		cmat_sk	Clave subrogada de la dimensión carrera_materia	dim_carrera_materia	cmat_sk	public
nombre_materia	character varying	300	cmat_nombre_materia	Campo donde se almacena el nombre de la materia	dim_carrera_materia		public
numero_creditos	integer		cmat_num_creditos	Campo donde se almacena el número de créditos de cada materia	dim_carrera_materia		public
id_carrera_materia	bigserial		cmat_id_carmateria	Campo donde se almacena el código de carrera_materia	dim_carrera_materia		public
nombre_carrera	character varying	300	cmat_nombre_carrera	Campo donde se almacena el nombre de la carrera	dim_carrera_materia		public
FECHA_ETL	date		aud_fecha	Campo donde registra la fecha de carga del ETL	dim_auditoria		sistema
NOMBRE_ETL	character varying		aud_nombreetl	Campo donde registra el nombre del proceso ETL	dim_auditoria		sistema
IP	character varying		aud_ip	Campo donde registra la IP de la maquina donde se creó el ETL	dim_auditoria		sistema

Tabla 4.11 Mapeo de información de la dimensión DIM_CARRERA_MATERIA

Fuente: Elaboración propia

Nombre de dimensión: DIM_TIEMPO

Tipo de tabla destino: dimensional

Esquema: Public

Descripción: Dimensión que conforma el cubo CARRERA_MATERIA, ALUMNO_MATERIA

Origen			Destino				
Campo	Tipo de dato	Tam	Nombre Columna	Descripción	Tabla	FK	Esquema
	bigserial		tie_sk	Clave subrogada de la dimensión tiempo	dim_tiempo	tie_sk	public
Fecha	date		tie_fecha	Campo que toma la fecha del sistema	dim_tiempo		public
Año	character varying	20	tie_anio	Campo que toma el año del sistema	dim_tiempo		public
Semestre	character varying	20	tie_semestre_ord	Campo que divide el año en semestres	dim_tiempo		public
Semestre largo	character varying	20	tie_semestre_l	Campo que divide el año en semestres y presenta la fecha en formato largo	dim_tiempo		public
Semestre corto	character varying	20	tie_semestre_c	Campo que divide el año en semestres y presenta la fecha en formato corto	dim_tiempo		public
Cuatrimestre	character varying	20	tie_cuatrimestre_ord	Campo que divide el año en cuatrimestres	dim_tiempo		public
Cuatrimestre largo	character varying	20	tie_cuatrimestre_l	Campo que divide el año en cuatrimestres y presenta la fecha en formato largo	dim_tiempo		public
Cuatrimestre corto	character varying	20	tie_cuatrimestre_c	Campo que divide el año en cuatrimestres y presenta la fecha en formato corto	dim_tiempo		public
Trimestre	character varying	20	tie_trimestre_ord	Campo que divide el año en trimestres	dim_tiempo		public
Trimestre largo	character varying	20	tie_trimestre_l	Campo que divide el año en trimestres y presenta la fecha en formato largo	dim_tiempo		public
Trimestre corto	character varying	20	tie_trimestre_c	Campo que divide el año en trimestres y presenta la fecha en formato corto	dim_tiempo		public
Mes	character varying	20	tie_mes_ord	Campo que toma el mes del sistema	dim_tiempo		public
Mes largo	character varying	20	tie_mes_l	Campo que toma el mes del sistema y presenta el mes en formato largo	dim_tiempo		public

Nombre de dimensión: DIM_TIEMPO (continuación)

Tipo de tabla destino: dimensional

Esquema: Public

Descripción: Dimensión que conforma el cubo CARRERA_MATERIA, ALUMNO_MATERIA

Origen			Destino				
Campo	Tipo de dato	Tam	Nombre Columna	Descripción	Tabla	FK	Esquema
Mes corto	character varying	20	tie_mes_c	Campo que toma el mes del sistema y presenta el mes en formato corto	dim_tiempo		public
Día semana corto	character varying	20	tie_dia_semana_c	Campo que toma el día del sistema y presenta el día en formato corto	dim_tiempo		public
Día semana largo	character varying	20	tie_dia_semana_l	Campo que toma el día del sistema y presenta el día en formato largo	dim_tiempo		public
Día año	character varying	20	tie_dia_anio_ord	Campo que toma el día del año del sistema	dim_tiempo		public
Día mes	character varying	20	tie_dia_mes_ord	Campo que toma el día del mes del sistema	dim_tiempo		public
Día semana	character varying	20	tie_dia_semana_ord	Campo que toma el día de la semana del sistema	dim_tiempo		public
FECHA_ETL	date		aud_fecha	Campo donde registra la fecha de carga del ETL	dim_auditoria		public
NOMBRE_ETL	character varying	300	aud_nombreetl	Campo donde registra el nombre del proceso ETL	dim_auditoria		public
IP	character varying	300	aud_ip	Campo donde registra la IP de la maquina donde se creó el ETL	dim_auditoria		public

Tabla 4.12 Mapeo de información de la dimensión DIM_TIEMPO

Fuente: Elaboración propia

- **Diseño de la base de datos del modelo dimensional para el cubo alumno_materia**

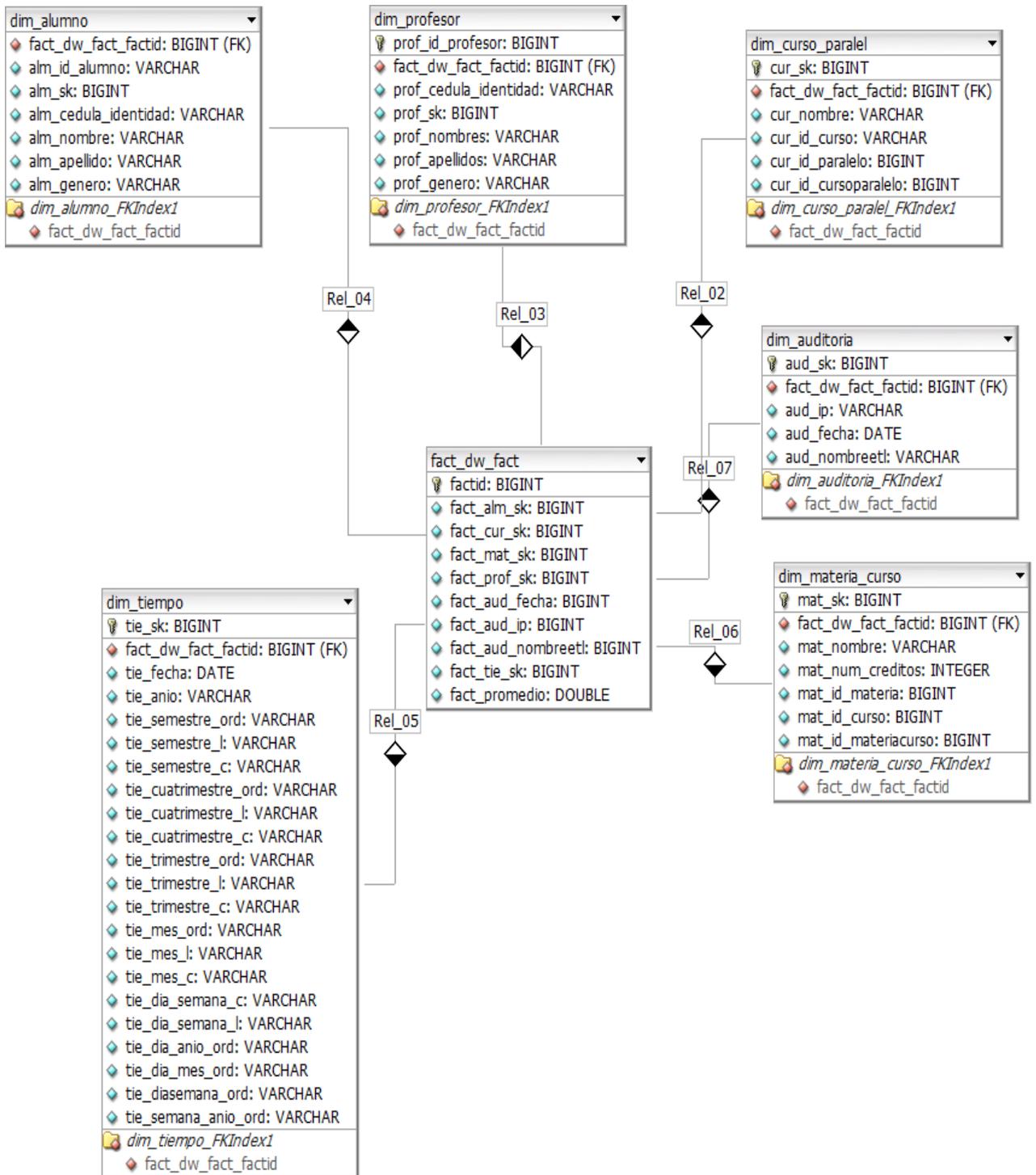


Figura 4.12 Base de datos dimensional alumno_materia
Fuente: Elaboración propia

- **Diseño de la base de datos del modelo dimensional para el cubo carrera_materia**

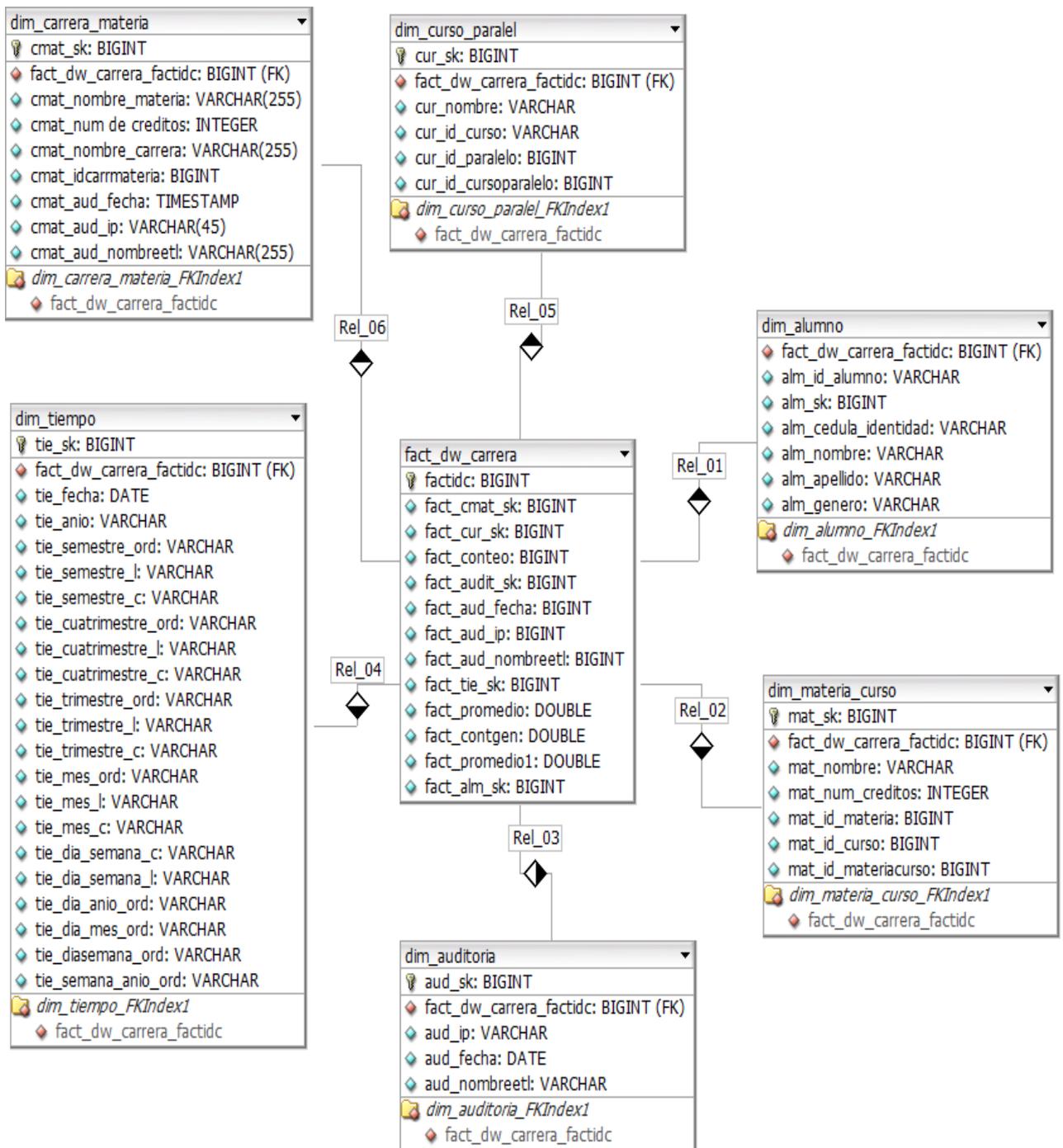


Figura 4.13 Base de datos dimensional carrera_materia
Fuente: Elaboración propia

4.3.4 Construcción del prototipo

Luego de haber seleccionado las herramientas que se van a utilizar para la construcción del prototipo, elaborado los modelos dimensionales, se procederá a elaborar el prototipo el cual funciona con la alimentación de información, provocando así una interacción entre el usuario y los equipos, con la finalidad de validar requerimientos importantes, para mantener las opciones de implementación; en esta etapa se describirá los pasos y configuraciones a seguir en el desarrollo del proyecto.

4.3.5 Desarrollo de dimensiones a través de una interfaz gráfica

Como ya se describió anteriormente la suite de Pentaho ofrece una herramienta muy eficiente para realizar el proceso ETL llamada Spoon que ofrece una interfaz amigable al usuario con la barra de las diferentes herramientas, a continuación se describirá más detalladamente el desarrollo del proceso ETL con esta herramienta, primero se empezará creando un nuevo trabajo o transformación, este proceso es similar para todas las dimensiones a generar.

4.3.5.1 TABLE INPUT O TABLA DE ENTRADA.- Esta herramienta permite leer información de las tablas de una base de datos transaccional, a través de la ejecución de una sentencia SQL, estos datos serán los que llenarán la base de datos dimensional.

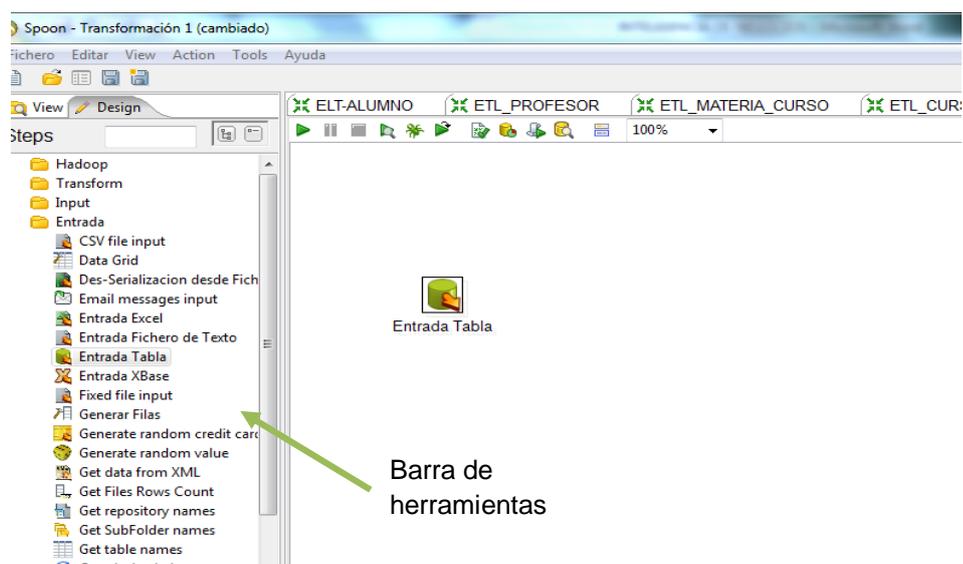


Figura 4.14 Insertando un Table Input
Fuente: Elaboración propia

Dar doble clic sobre el ícono Tabla de Entrada se despliega una ventana donde se harán las siguientes configuraciones:

- Nombre de la Tabla de Entrada.
- Sentencia SQL para extraer los datos.
- Conexión con la base de datos.

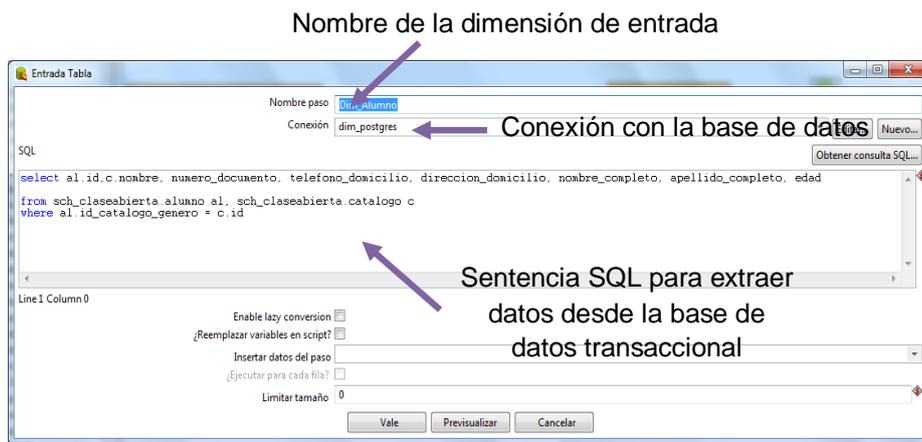


Figura 4.15 Configuración de un Table Input
Fuente: Elaboración propia

Ahora se configurará la conexión con la base de datos, para ello se ubicará en nuevo y a continuación se desplegará la siguiente pantalla:

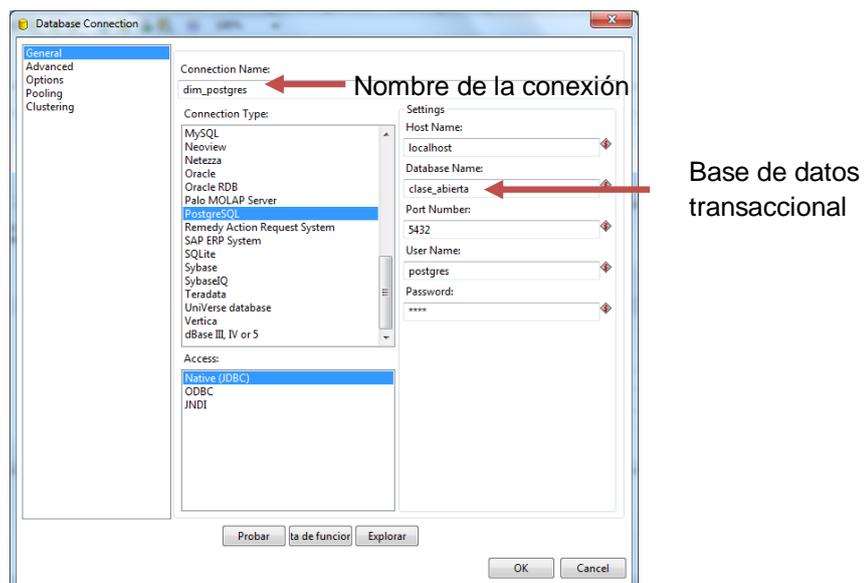


Figura 4.16 Configuración de conexión de base de datos de un Table Input
Fuente: Elaboración propia

Hay algunos parámetros que son configurables por ejemplo:

- **Host Name:** aquí se configura la IP del servidor con la que va a trabajar la base de datos.
- **Database Name:** Nombre de la base de datos transaccional.
- **Port Name:** Numero de puerto que utiliza cada una de las base de datos de la lista.
- **User Name:** Nombre del usuario de la base de datos.
- **Password:** Contraseña con la que accede el usuario a la base de datos.

Luego de haber creado la conexión con la base de datos se prueba si está bien hecha la conexión:

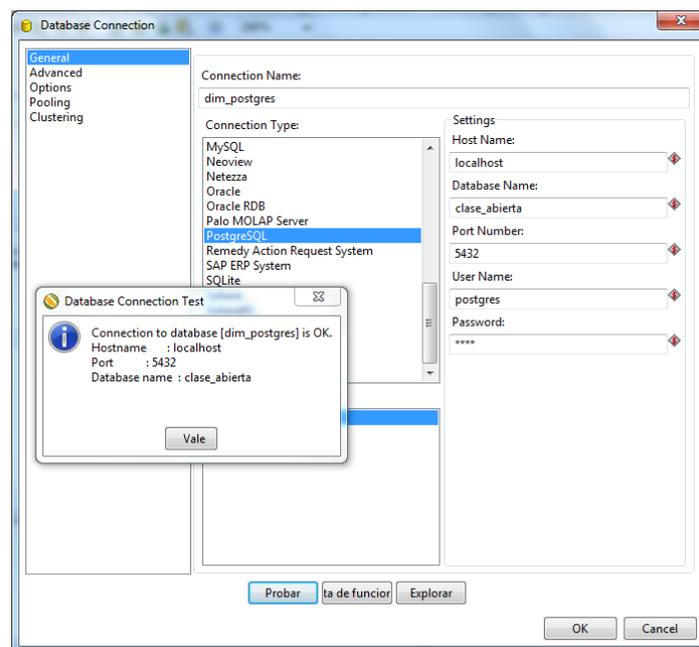


Figura 4.17 Probando conexión de base de datos
Fuente: Elaboración propia

Dar clic en **VALE** para verificar la correcta conexión con la base de datos y luego dar clic en **OK** para terminar la configuración de la conexión, si se desea cambiar algún parámetro antes mencionado solo debe dar clic en el botón **EDITAR**. En el botón **PREVISUALIZAR** podrá ver la información que se ha seleccionado a través del query SQL.

#	id	nombre	numero_documento	telefono_domicilio	direccion_domicilio
1	12	MASCULINO	12	2240634/2228424	Zamora y Edmundo Chiriboga
2	32	MASCULINO	1272095844	2569980	NAVARRO n 23-85 y LA GASCA (NAVARRO N23-85 Y LA GASCA)
3	22	MASCULINO	22	2439330	PABLO CASALS 46-24 y LAS RETAMAS (JUNTO A CANCHA DE BASQUET)
4	25	FEMENINO	25	2502846	FERNANDEZ DE RECALDE N23-136 y DIEGO ZORILLA FERNANDEZ DE RECALDE N23-136)
5	26	MASCULINO	26	3215504	EUSTORGIO SALGADO n23-223 y ARMERO (EUSTORGIO SALGADO N23-202)
6	4	FEMENINO	1753163953	2554183	GASPAR DE CARVAJAL N38-44 y CUERO Y CAicedo (DOS CUADRAS AL SUR DEL COLEGIO SAN GABRIEL)
7	14	MASCULINO	1727092346	6012411	PAISO ea5-374 y PAISO DEL SOL (CANTAROS III CASA 10)
8	39	FEMENINO	1726054339	2508 807	DOMINGO ESPINAR n23-183 y LA GASCA (DOMINGO ESPINAR 23-183 Y LA GASCA)
9	52	MASCULINO	1725562988	2548239	RUIRANCA ES72 y POLONA (RUIRANCA ES72 Y POLONA)
10	61	FEMENINO	61	2959256	BRACAMONTE OE13-04 y ALVARO DE CEVALLOS (BRACAMONTE OE13-04)
11	65	FEMENINO	65	2241016	Fcco. Miranda N 48-296
12	37	MASCULINO	1726480452	2404610	PASAJE C m y SAN JOSE MARIA (CONI FLOR DEL PARAISO CASA 11)
13	2	FEMENINO	49763054	098622463	COLON oa3-21 y VERSALLES (EDIFICIO VILLAREAL)
14	70	FEMENINO	70	3004140	Tiguarcano # 200 y Cerro Namo
15	75	FEMENINO	75	2477792	Cinuelos y Tulipanes
16	79	MASCULINO	79	2650146	Maldonado y Vicente Andrade
17	56	MASCULINO	56	2222858	ACEVEDO 24 45 y LA GASCA (JUNTO AL TENNIS ECUADOR)
18	43	MASCULINO	1752512991	02209628	SAN PEDRO n23-22 y RUMPA MBA (UNIVERSIDAD EQUINOCCIAL)
19	80	MASCULINO	172600026-6	2529674	LAS GUAYANAS 323 y 10 DE AGOSTO (GUAYANAS 323 Y AV. 10 AGOSTO)
20	86	MASCULINO	86	2677721	BORBON ea2-158 y PASI OE2-C (TURUBAMBA BAJO SM4, C11, REGISTRO CIVIL)
21	89	FEMENINO	89	2619580	Salvador Bravo y Av. Tnte. Hugo Ortiz
22	203	MASCULINO	1725456048	6040313	GONZALEZ SUAREZ e11-193 y Av. BALO (GONZALEZ SUAREZ E BALO ESQ. VALLE DE LOS CHILLOS)
23	141	MASCULINO	141	2477792	Los Cinuelos 186-238 y Tulipanes
24	64	MASCULINO	1727095836	2569980	NAVARRO n23-85 y LA GASCA (NAVARRO N23-85 Y LA GASCA)
25	297	FEMENINO	297	2241016	Fcco. Miranda N 48-296 Pinar Bajo
26	68	FEMENINO	68	3026764	TABIAZO 0e-1105 y PROFETA MIQUEAS (PANADERIA)
27	71	MASCULINO	1729915944	2509572	ATACAMES n23-276 y LA GASCA (ATACAMES N23-276 Y LA GASCA)
28	77	FEMENINO	1725566747	3902142	AMAZONAS OE187 y JUAN HOLGUIN FRENTE A LA ENTRADA PARQUE BICENTENARIO)
29	53	FEMENINO	172587796-4	2840452	RAFAEL RAMOS 43-79 y FRANCISCO SALGADO (RAFAEL RAMOS 43-79 Y FCO. SALGADO)
30	59	MASCULINO	59	5104092	JUAN DE LA MERCED N83-06 y ALONSO DIAZ (JUAN DE LA MERCED N83-06 Y ALONSO DIAZ)
31	94	MASCULINO	94	3200030	Calle Antonio Herrera Lote 9 y José Bernueta La Gasca

Figura 4.18 Visualización de datos cargados en un Table Input
Fuente: Elaboración propia

4.3.5.2 ADD SEQUENCE O AGREGAR SECUENCIA.- Al insertar esta herramienta nos ayuda a crear secuencias en la base de datos dimensional del Datamart.

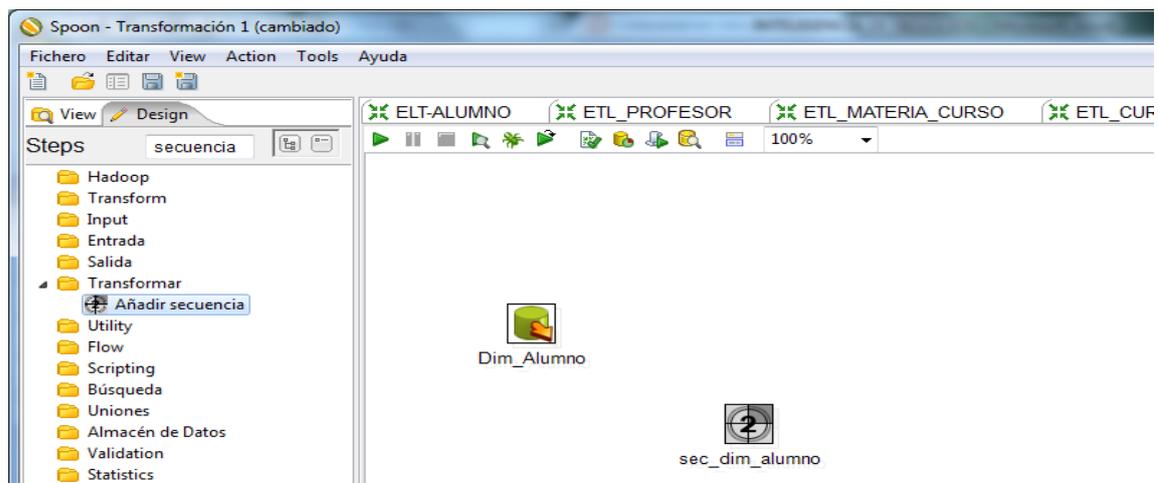


Figura 4.19 Agregando una Secuencia
Fuente: Elaboración propia

Al igual que las demás herramientas necesita ser configurada, así que al dar doble clic, los parámetros configurables son los siguientes:

- **Nombre de paso:** O nombre de secuencia, la misma que va a ser usada para conectar el datamart.
- **Nombre de valor:** Clave primaria para relacionar la base de datos dimensional con el datamart.

A continuación, se marca la opción **UTILIZAR BASE DE DATOS PARA OBTENER SECUENCIA** y se configuran las opciones de una nueva **Conexión** con la base de datos dimensional, los campos a configurar son:

- **Connection name:** Nombre de la nueva conexión con la base de datos dimensional.
- **Host Name:** Aquí se configura la IP del servidor con la que va a trabajar la base de datos.
- **Database Name:** Nombre de la nueva base de datos dimensional.
- **Port Name:** Numero de puerto que utiliza cada una de las base de datos de la lista.
- **User Name:** Nombre del usuario de la base de datos.
- **Password:** Contraseña con la que accede el usuario a la base de datos.

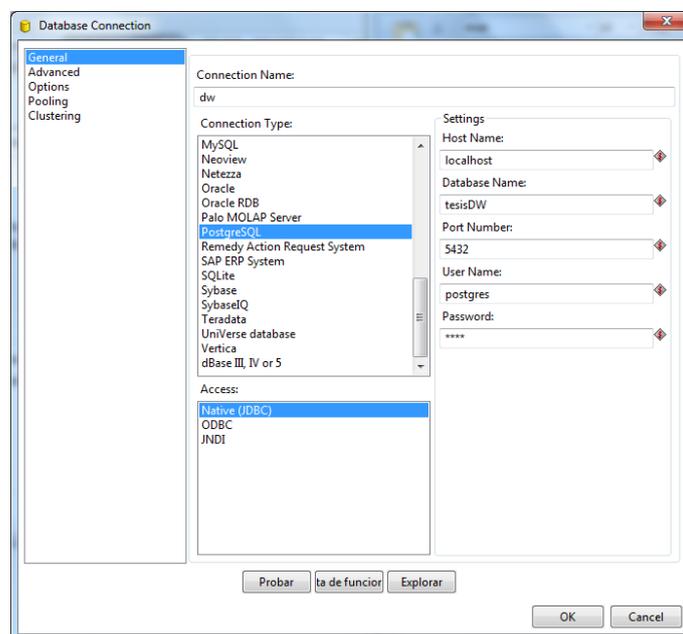


Figura 4.20 Agregando una Nueva Conexión con la Base de Datos Dimensional
Fuente: Elaboración propia

Dar clic en **VALE** al verificar la correcta conexión con la base de datos y dar clic en **OK** para terminar la configuración de la conexión, si se desea cambiar algún parámetro antes mencionado solo debe dar clic en el botón **EDITAR**.

Luego de haber configurado se prueba la nueva conexión.

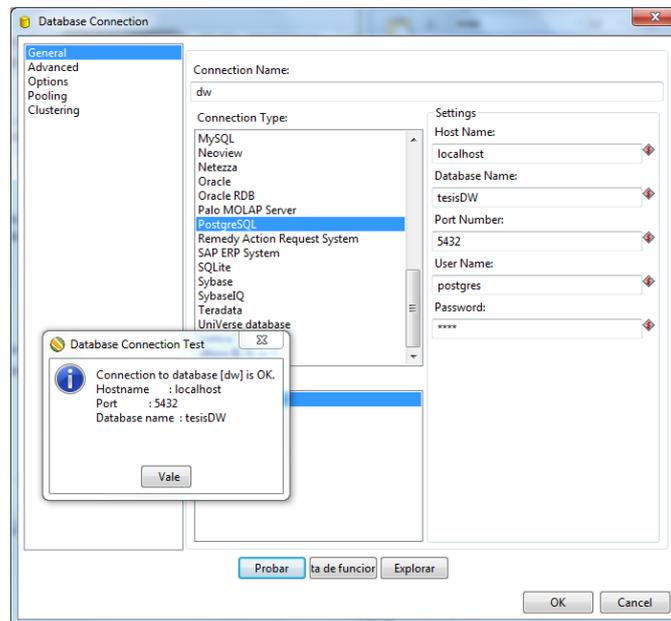


Figura 4.21 Probando conexión con la base de datos dimensional
Fuente: Elaboración propia

- **Nombre de esquema:** Esquema en el que se encuentra creada la base de datos dimensional.
- **Nombre de secuencia:** Secuencia que se crea al momento de crear las tablas dimensionales de la base de datos dimensional.

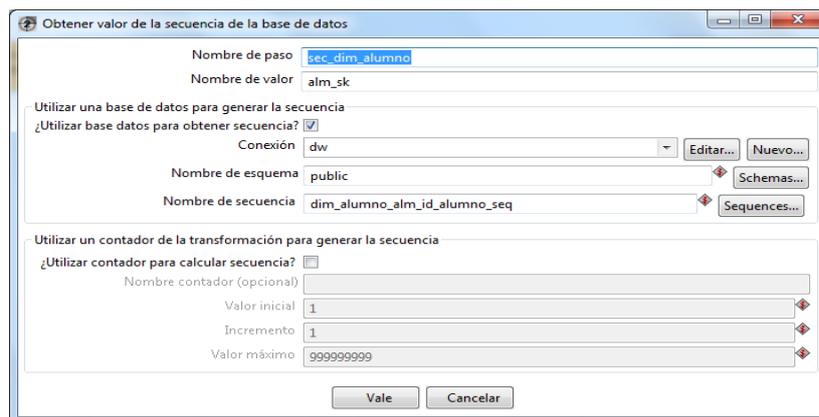


Figura 4.22 Configuración de Secuencia de una nueva dimensión
Fuente: Elaboración propia

4.3.5.3 GET SYSTEM INFO O OBTENER INFORMACIÓN DEL SISTEMA.- Esta herramienta permite obtener datos propios del sistema, mismos que serán utilizados en la tabla de auditoría.

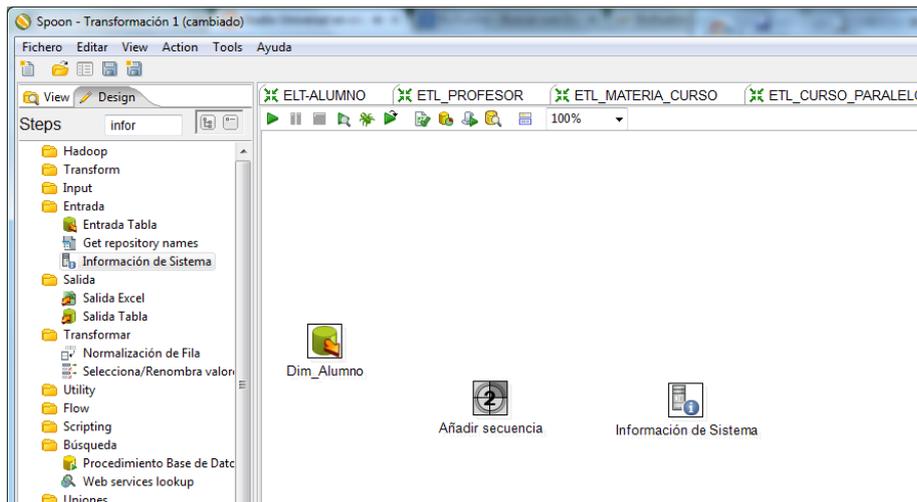


Figura 4.23 Insertando Get System Info
Fuente: Elaboración propia

A continuación al dar doble clic se agregarán y configurarán los siguientes campos:

- **Fecha de ETL:** Toma automáticamente la fecha de la maquina donde se está elaborando y configurando las dimensiones del Datamart.
- **Nombre ETL:** Aquí toma un nombre fijo del sistema.
- **IP:** Toma automáticamente la dirección IP de la maquina donde se está elaborando y configurando las dimensiones del Datamart.

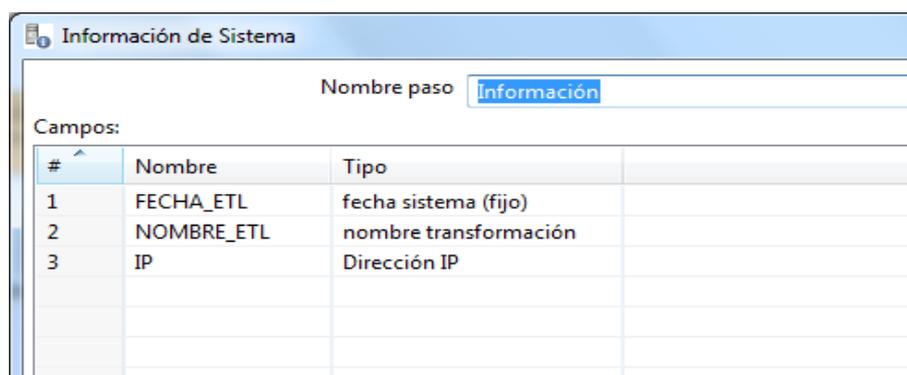


Figura 4.24 Configuración de Get System Info
Fuente: Elaboración propia

4.3.5.4 INSERT/UPDATE O INSERCIÓN / ACTUALIZACIÓN.- Esta Herramienta permite controlar de manera ordenada la inserción o actualización de la información que va a ser ingresada en la base de datos dimensional.

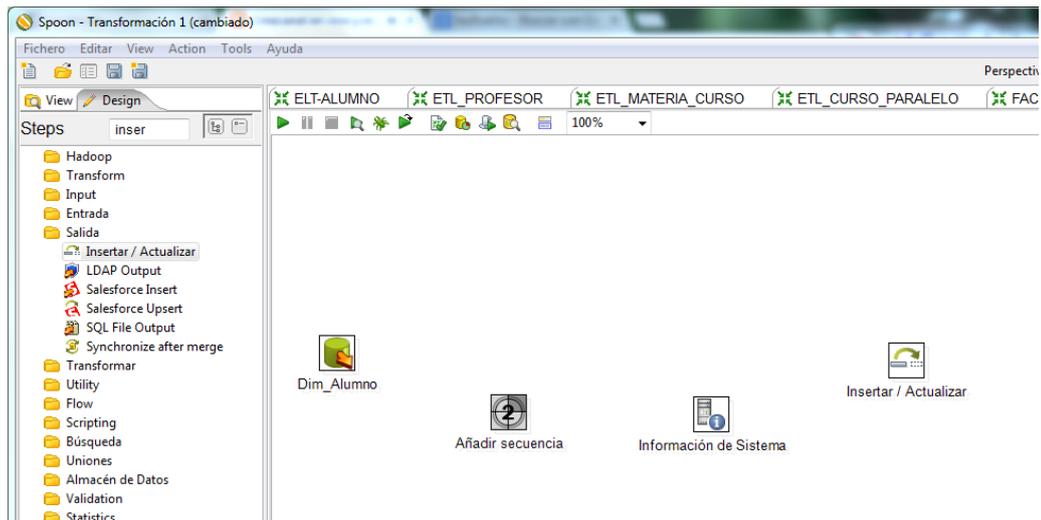


Figura 4.25 Insertando un Insert/Update
Fuente: Elaboración propia

Al momento de dar doble clic, se despliega una nueva ventana, a continuación se describirá los campos a configurar en esta herramienta, cabe mencionar que esta herramienta consta de tres bloques de configuración: la parte de conexión con la base de datos, la parte de configuración de claves y la parte de configuración de campos de actualización, a continuación se describirá a cada uno de ellos:

- **Nombre de paso:** Aquí se asigna un nombre para identificar el proceso.

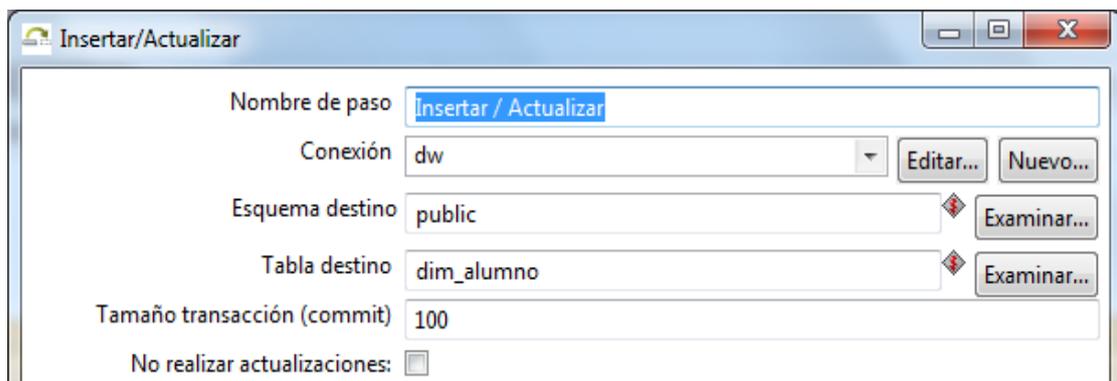


Figura 4.26 Configuración herramienta de Inserción / Actualización
Fuente: Elaboración propia

- **Conexión:** Nombre de la conexión que se va a conectar con la base de datos dimensional.

A continuación, dar clic en nuevo y se configurarán las opciones de una nueva **Conexión** con la base de datos dimensional, los campos a configurar son:

- **Connection name:** Nombre de la nueva conexión con la base de datos dimensional.
- **Host Name:** Aquí se configura la IP del servidor con la que va a trabajar la base de datos.
- **Database Name:** Nombre de la nueva base de datos dimensional.
- **Port Name:** Numero de puerto que utiliza cada una de las base de datos de la lista.
- **User Name:** Nombre del usuario de la base de datos.
- **Password:** Contraseña con la que accede el usuario a la base de datos.

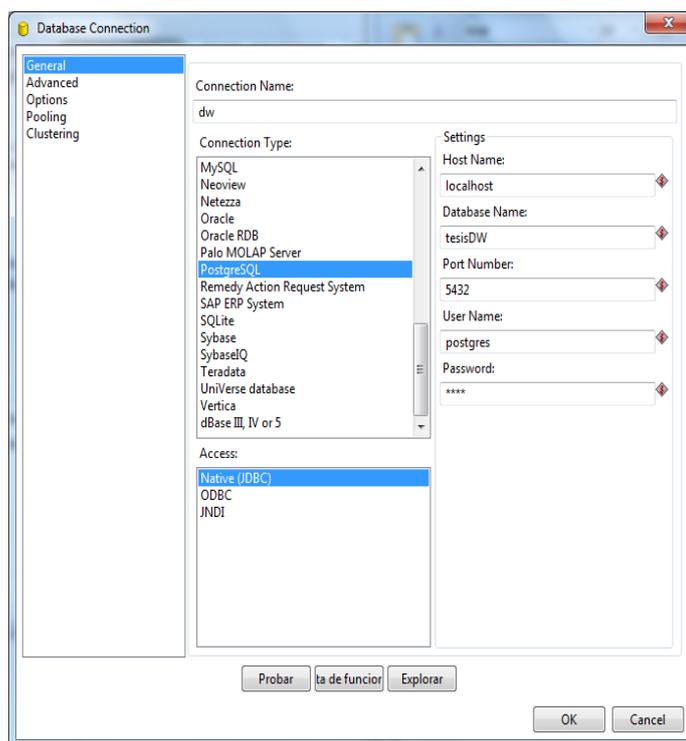


Figura 4.27 Agregando una Nueva Conexión con la Base de Datos Dimensional
Fuente: Elaboración propia

Dar clic en **VALE** al verificar la correcta conexión con la base de datos y dar clic en **OK** para terminar la configuración de la conexión, si se desea cambiar algún parámetro antes mencionado solo dar clic en el botón **EDITAR**.

Luego de haber configurado se prueba la nueva conexión.

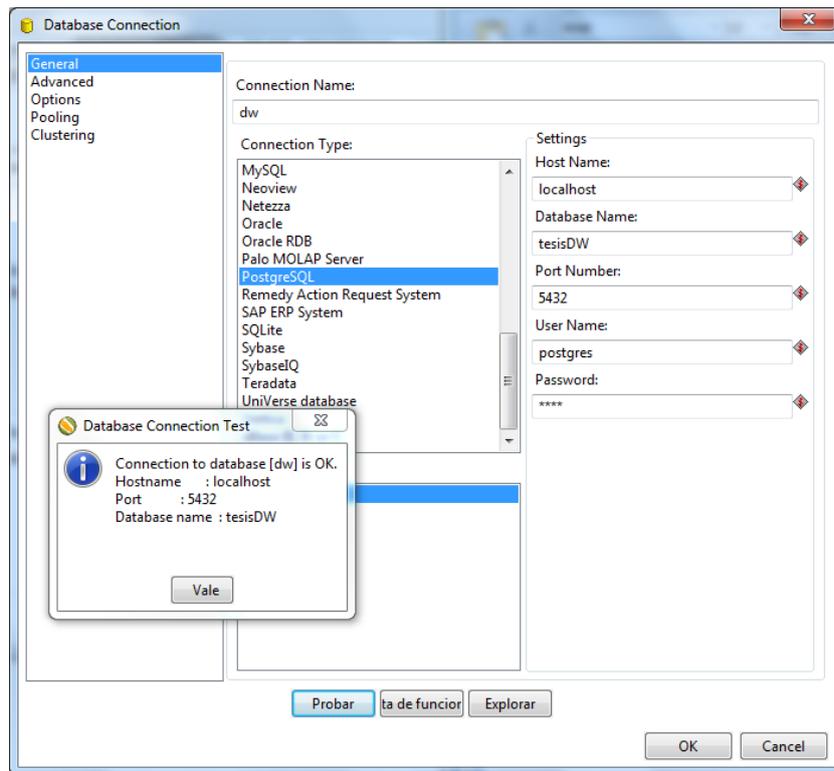


Figura 4.28 Probando conexión con la base de datos dimensional
Fuente: Elaboración propia

- **Esquema destino:** Aquí se configura el nombre del esquema de la nueva tabla que va almacenar la información en la base de datos dimensional.
- **Tabla destino:** Es el nombre de la tabla dimensional que va almacenar la nueva información dentro de la nueva base de datos que conformará el datamart.

En la sección que contiene la parte de configuración de **LAS CLAVES PARA REALIZAR LA BÚSQUEDA DE VALORES** se seleccionará las claves primarias o ID tanto de la base de datos transaccional que sería la base de datos de origen como de la base de datos dimensional que sería la base del datamart, es decir hacer un JOIN entre ambas ID's.

En el **CAMPO DE TABLA** se seleccionará el campo que contiene la clave primaria o ID de la nueva tabla dimensional. En el campo **COMPARADOR** seleccionar un operador lógico que permite comparar ambas claves primarias.

En el **CAMPO1** seleccionar el campo que contiene la clave primaria o ID de la tabla de la base de datos transaccional.

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

#	Campo de tabla	Comparador	Campo1	Campo2
1	alm_id_alumno	=	id	

Obtener Campos

Figura 4.29 Configuración de claves primarias
Fuente: Elaboración propia

Continuando con la configuración en la sección que contiene la parte de configuración de **CAMPOS DE ACTUALIZACIÓN** se seleccionarán y se relacionarán los campos de la base de datos transaccional con los nuevos campos de la base de datos dimensional.

Campos de actualización:

#	Campo de tabla	Campo de Flujo	Actuali
1	alm_edad	edad	Y
2	alm_cedula_identidad	numero_documento	Y
3	alm_telefono	telefono_domicilio	Y
4	alm_direccion	direccion_domicilio	Y
5	alm_nombre	nombre_completo	Y
6	alm_apellido	apellido_completo	Y
7	alm_sk	alm_sk	Y
8	alm_genero	nombre	Y
9	alm_id_alumno_1	id	Y

Obtener campos de actualización

Edit mapping

Vale Cancelar SQL

Figura 4.30 Configuración de campos de actualización
Fuente: Elaboración propia

Dar clic en **EDIT MAPPING**, se desplegará una nueva ventana donde podrá relacionar los campos tanto de la base de datos de origen como con la base de datos destino.

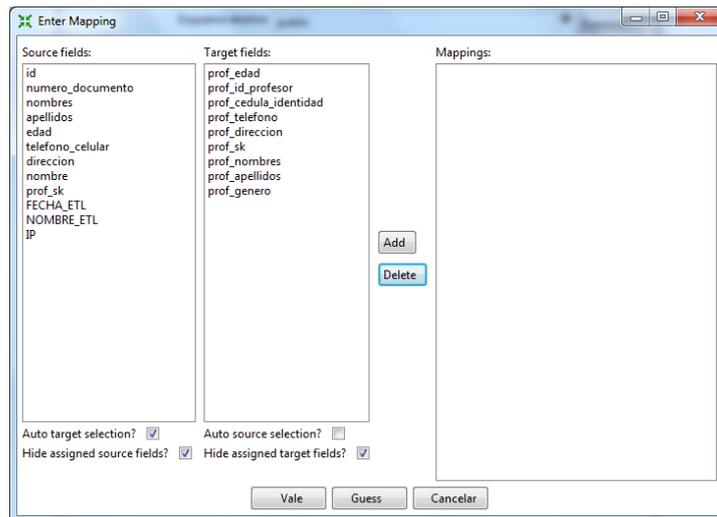


Figura 4.31 Ventana de Mapping de campos de relación
Fuente: Elaboración propia

Al momento de dar clic en el campo de origen, automáticamente se marca con el respectivo campo de destino, dar clic en ADD para que se agregue al mapa de relación de campos.

Campos de base de datos origen

Campos de base de datos destino

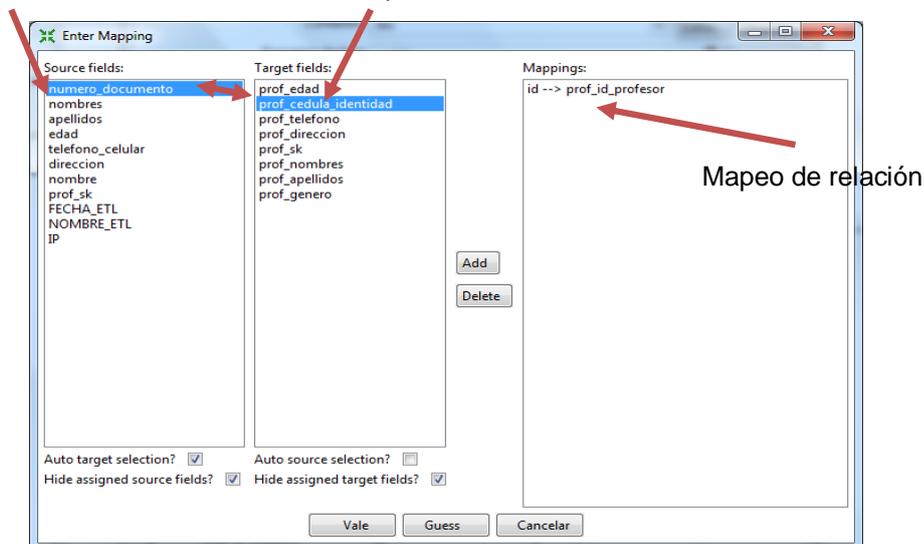


Figura 4.32 Configuración de Mapping de campos de relación
Fuente: Elaboración propia

Una vez terminado el proceso de relacionamiento de campos se puede observar el área de mapping de los campos relacionados, dar clic en **VALE** para cerrar la ventana de configuración.

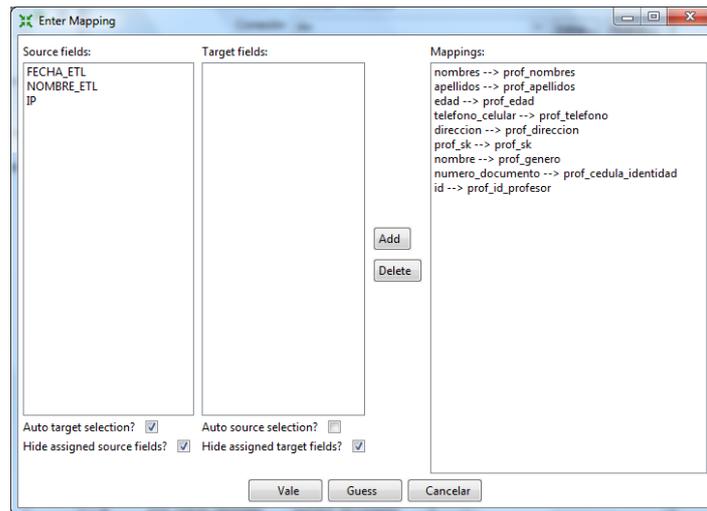


Figura 4.33 Configuración de Mappings
Fuente: Elaboración propia

Terminado todo el proceso de configuración podrá acceder a la opción de ver las sentencias SQL que realiza internamente la herramienta.

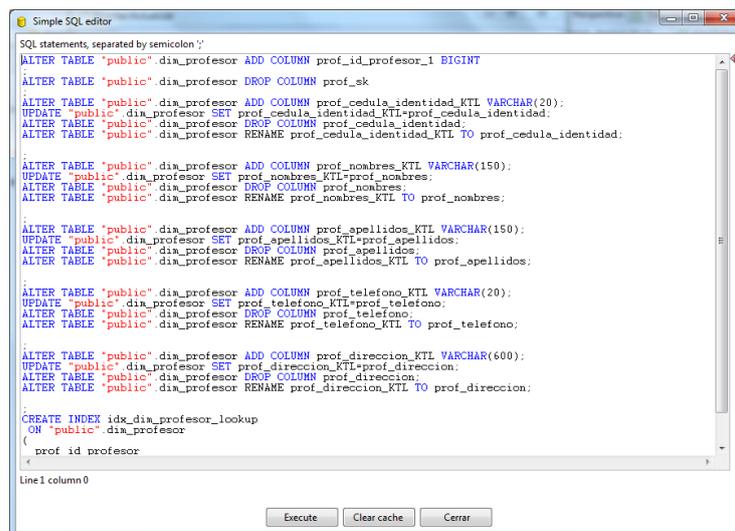


Figura 4.34 Sentencias SQL que ejecuta la herramienta Insertar/Actualizar internamente
Fuente: Elaboración propia

Se deberá agregar una nueva herramienta de secuencia de errores para finalizar el proceso de ETL de cada una de las dimensiones propuestas que conformarán el datamart.

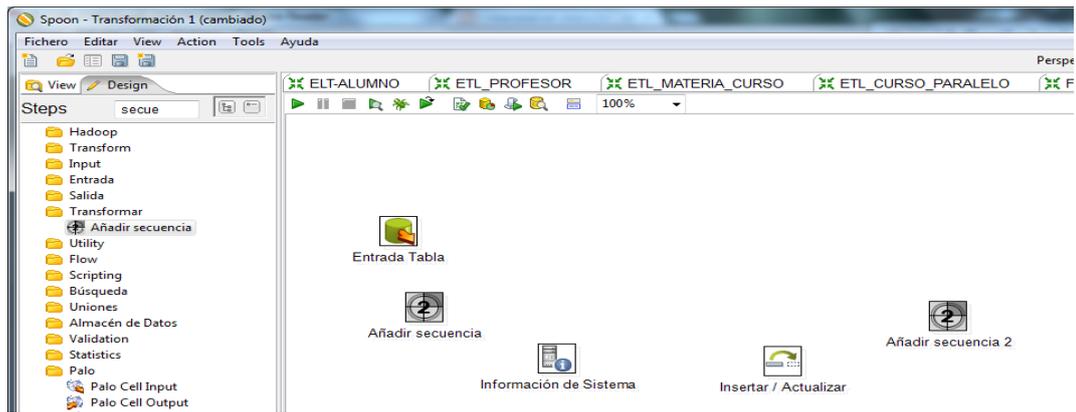


Figura 4.35 Agregando una Secuencia para errores
Fuente: Elaboración propia

Al igual que las demás herramientas necesita ser configurada, así que al dar doble clic, los parámetros configurables son los siguientes:

- **Nombre de paso:** O nombre de secuencia, la misma que va a ser usada para conectar el datamart.
- **Nombre de valor:** Clave primaria para relacionar la base de datos dimensional con el datamart.
- **Conexión:** Configuración de la conexión con la base de datos dimensional.
- **Nombre de esquema:** Esquema en el que se encuentra creada la base de datos dimensional.
- **Nombre de secuencia:** Secuencia que se crea al momento de crear las tablas dimensionales de la base de datos dimensional.

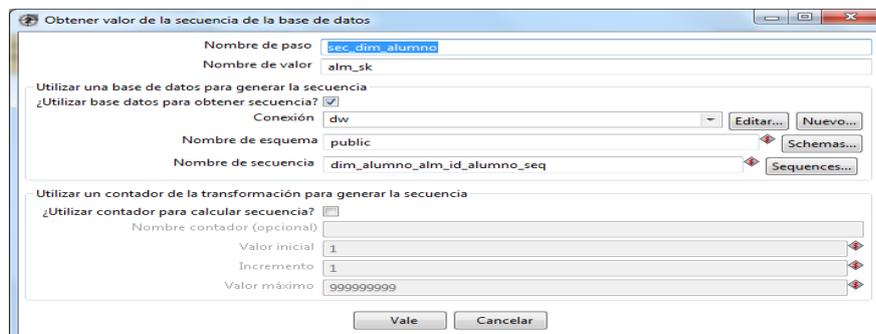


Figura 4.36 Configuración de Secuencia de una nueva dimensión
Fuente: Elaboración propia

Al momento de dar clic en **VALE** se termina la configuración, a continuación se describirá el proceso de configuración de la secuencia de errores para la tabla de auditoría.

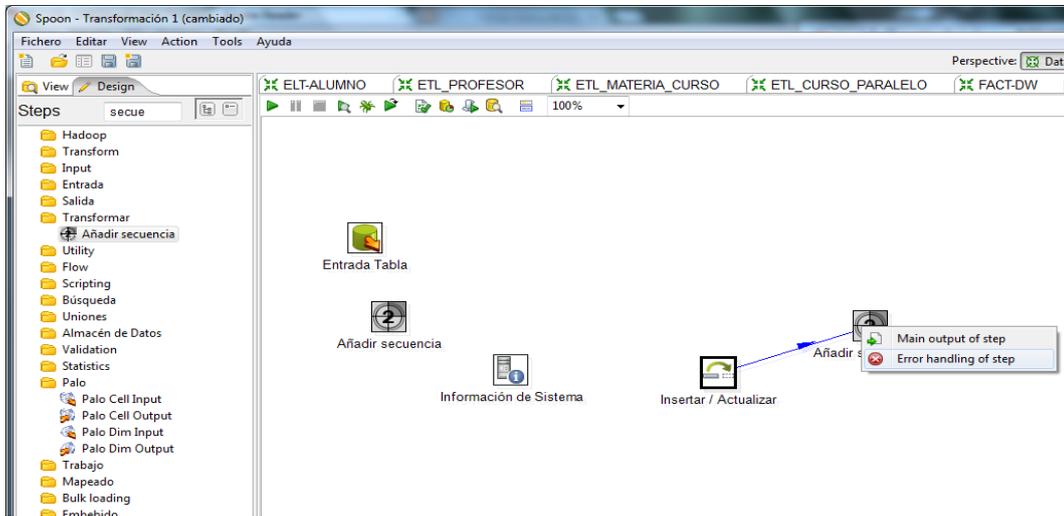


Figura 4.37 Unión de las herramientas Insertar/Actualizar con Secuencia
Fuente: Elaboración propia

Al momento de unir ambas herramientas podrá escoger dos opciones, para el caso se escogerá la opción **ERROR HANDLING OF STEP** es la que permitirá controlar los campos de error.

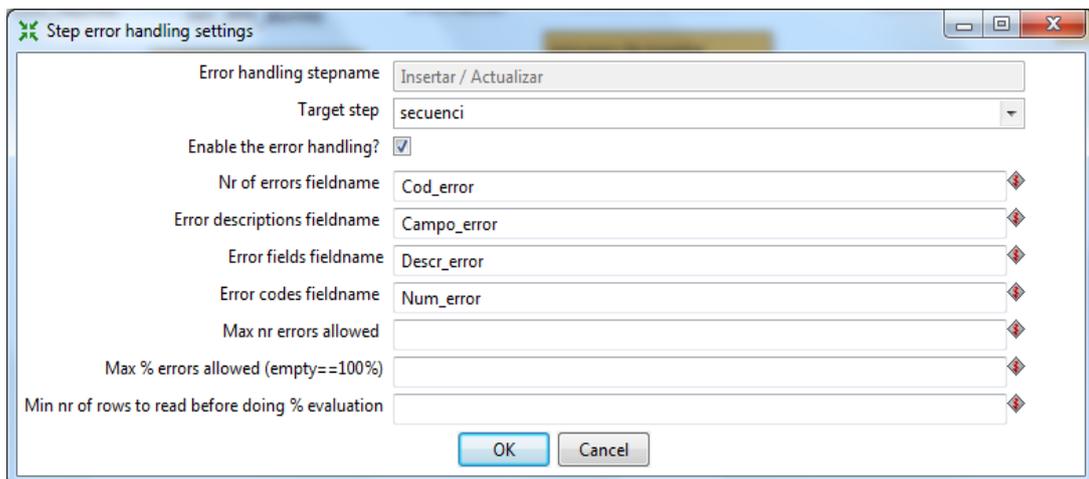


Figura 4.38 Configuración del Error Handling Step
Fuente: Elaboración propia

A continuación se configurarán los siguientes campos:

- **Target Step o Nombre de Etapa:** Nombre de secuencia, la misma que va a ser usada para conectar el datamart.
- **Nr of errors fieldname:** O número de errores de nombre de campo.
- **Error descriptions fieldname:** O nombre del campo que contiene la descripción del error.
- **Error fields fieldname:** O campos de error.
- **Error codes fieldname:** O los códigos de error de nombre de campo.

4.3.5.5 TABLE OUTPUT O TABLA DE SALIDA.- Esta es la última herramienta que se agrega en el proceso de ETL, aquí se almacenarán los errores y los campos de auditoría que se definieron anteriormente.

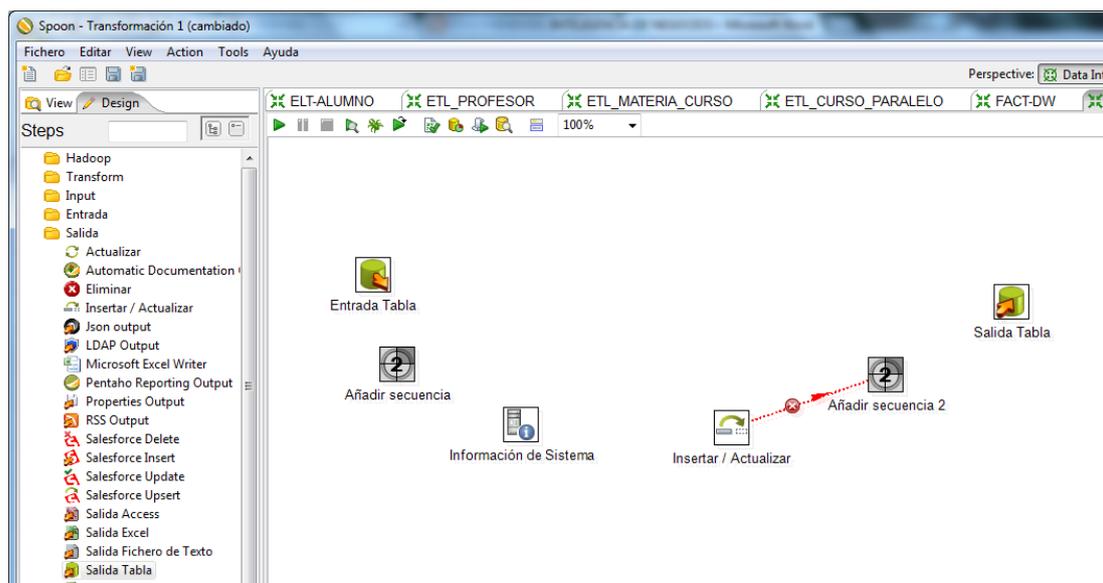


Figura 4.39 Agregando un Table Output
Fuente: Elaboración propia

Al momento de dar doble clic, se desplegará una nueva ventana, a continuación se describirá los campos a configurar en esta herramienta, esta herramienta consta de dos bloques de configuración: la parte de conexión con la base de datos, y la parte de configuración de los campos base o del sistema, a continuación se describirá a cada uno de ellos:

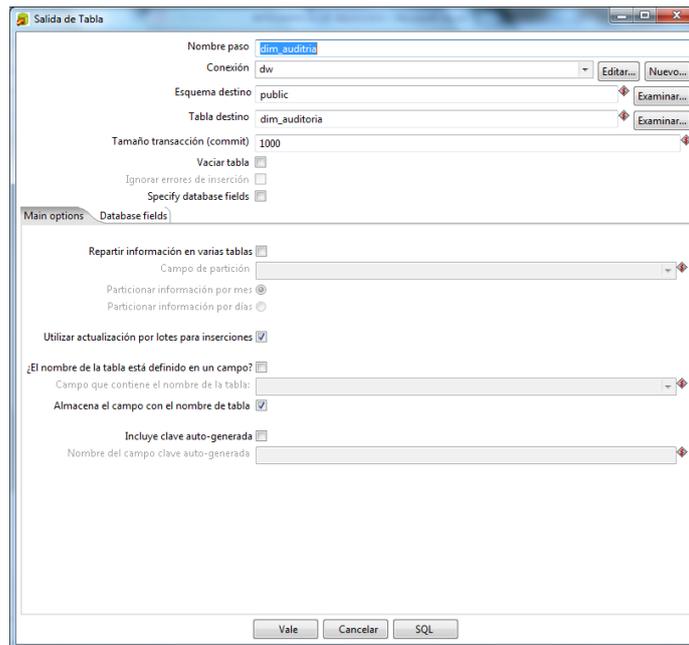


Figura 4.40 Configuración del Table Output
Fuente: Elaboración propia

- **Nombre de paso:** Nombre de la herramienta, la misma que será usada para conectar el datamart.
- **Conexión:** Configuración de la conexión con la base de datos dimensional.

A continuación, dar clic en nuevo y se configurarán las opciones de una nueva **Conexión** con la base de datos dimensional, los campos a configurar son:

- **Connection name:** Nombre de la nueva conexión con la base de datos dimensional.
- **Host Name:** Aquí se configura la IP del servidor con la que va a trabajar la base de datos.
- **Database Name:** Nombre de la nueva base de datos dimensional.
- **Port Name:** Numero de puerto que utiliza cada una de las base de datos de la lista.
- **User Name:** Nombre del usuario de la base de datos.
- **Password:** Contraseña con la que accede el usuario a la base de datos.

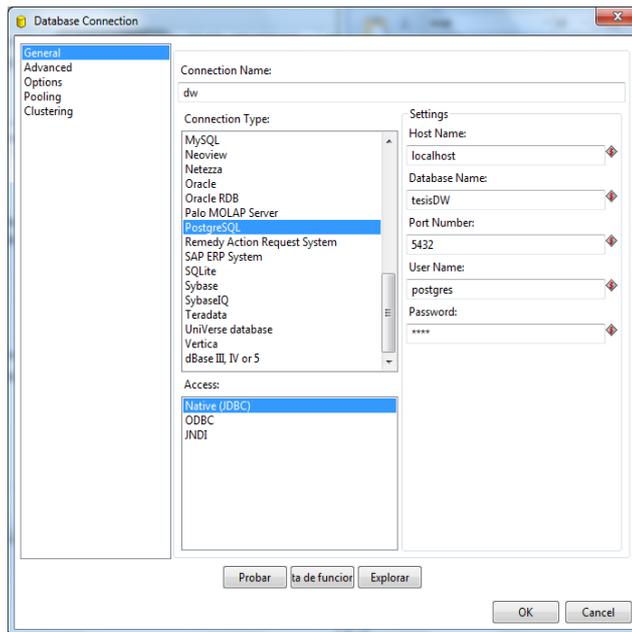


Figura 4.41 Agregando una Nueva Conexión con la Base de Datos Dimensional
Fuente: Elaboración propia

Dar clic en **VALE** al verificar la correcta conexión con la base de datos y dar clic en **OK** para terminar la configuración de la conexión, si desea cambiar algún parámetro antes mencionado solo dar clic en el botón **EDITAR**.

Luego de haber configurado se prueba la nueva conexión.

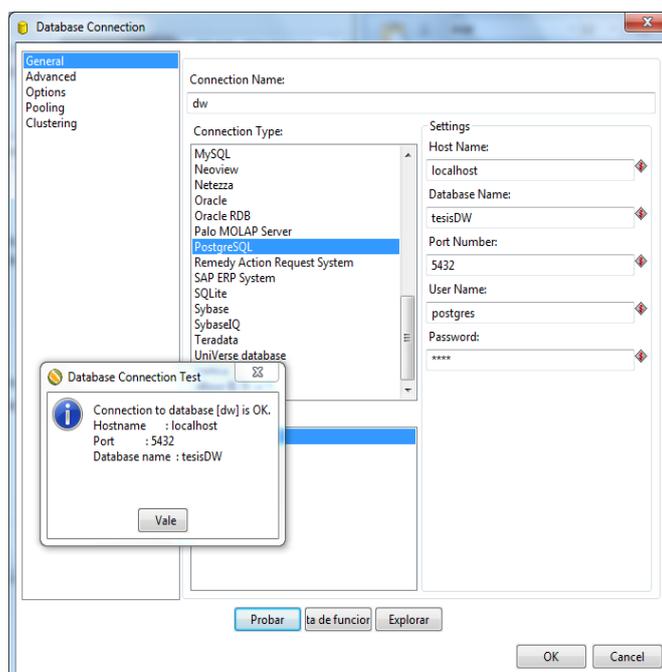


Figura 4.42 Probando conexión con la base de datos dimensional
Fuente: Elaboración propia

- **Esquema destino:** Aquí se configura el nombre del esquema de la nueva tabla que va almacenar la información en la base de datos dimensional.
- **Tabla destino:** Es el nombre de la tabla dimensional que va almacenar la nueva información dentro de la nueva base de datos que conformará el datamart.

Ahora se ubicará en la pestaña **DATABASE FIELDS** para realizar la respectiva configuración de los campos de la base de datos de origen con los datos de la base de datos destino.



Figura 4.43 Configurando los campos del Database Fields
Fuente: Elaboración propia

Dar clic en **ENTER FIELD MAPPING**, se desplegará una nueva ventana donde tendrá que relacionar los campos tanto de la base de datos de origen como con la base de datos destino.

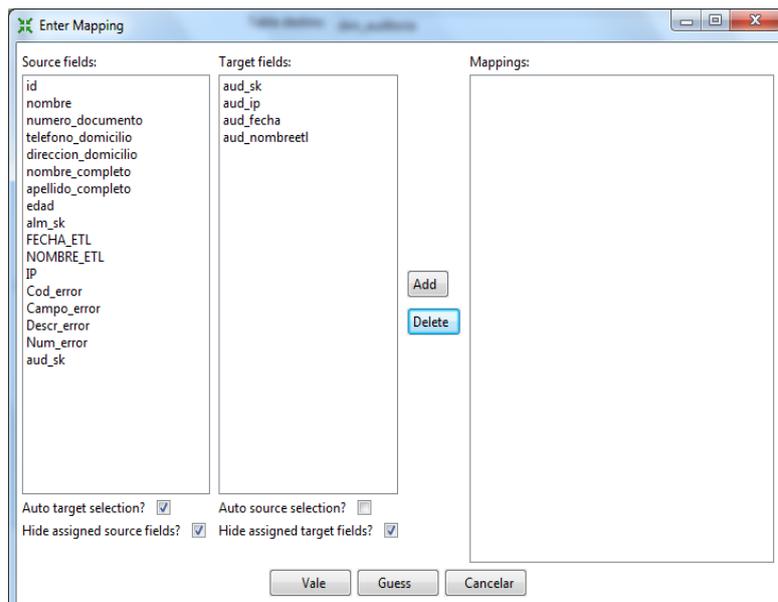


Figura 4.44 Ventana de Mapping de campos de relación
Fuente: Elaboración propia

Al momento de dar clic en el campo de origen, automáticamente se marca con el respectivo campo de destino, dar clic en **ADD** para que se agregue al mapa de relación de campos.

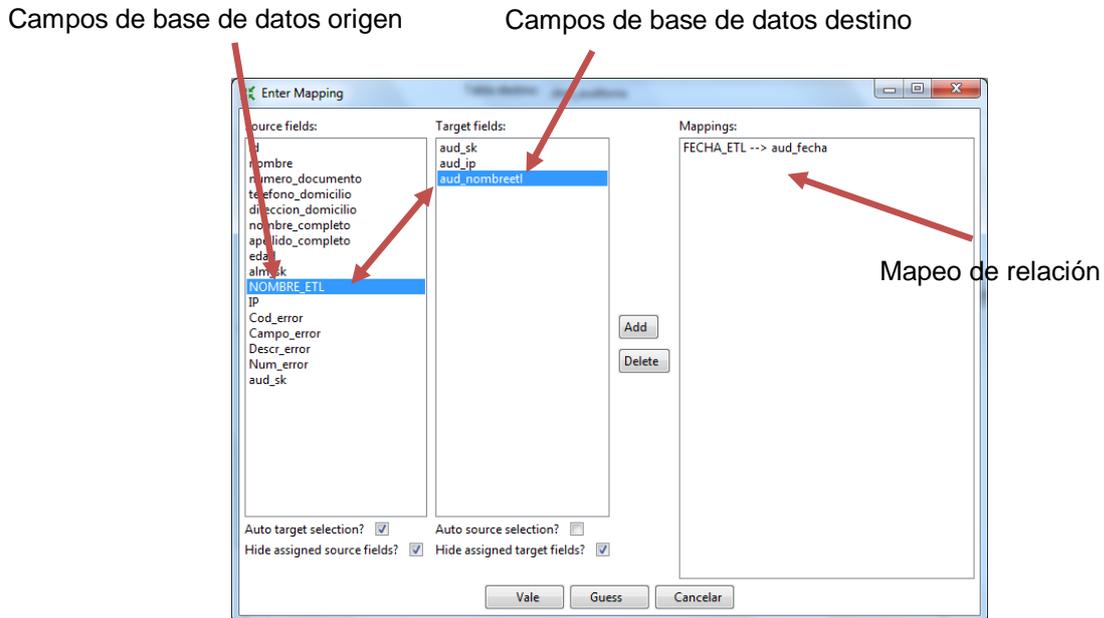


Figura 4.45 Configuración de campos de relación para la tabla de auditoría
Fuente: Elaboración propia

Una vez terminado el proceso de relacionamiento de campos podrá darse cuenta en el área de mapping que los campos relacionados, dar clic en **VALE** para cerrar la ventana de configuración.

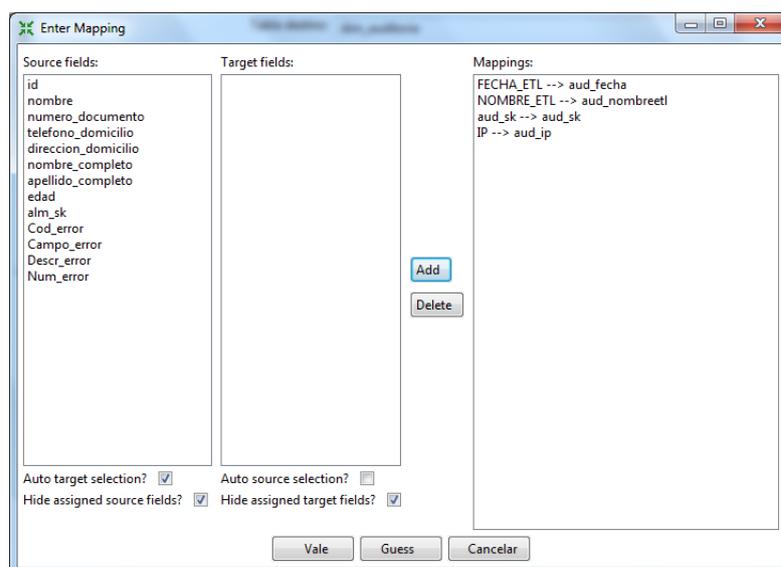


Figura 4.46 Configuración de Mappings
Fuente: Elaboración propia

Terminado todo el proceso de configuración tendrá la opción de ver las sentencias SQL que realiza internamente la herramienta.

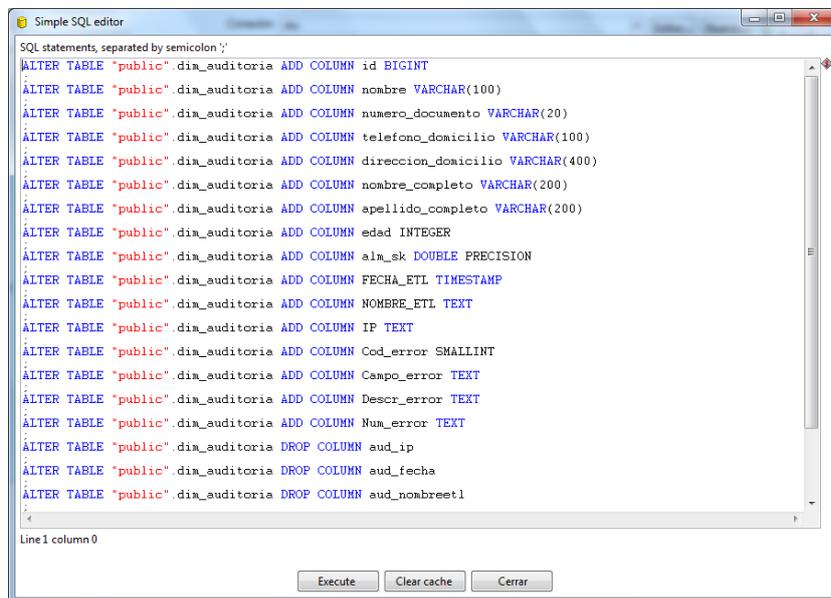


Figura 4.47 Sentencias SQL que ejecuta la herramienta Table Output internamente
Fuente: Elaboración propia

En la siguiente figura se muestra como todas las herramientas para diseño de un ETL se unen en un solo proceso solo es necesario ejecutar y comprobar si está bien realizado este proceso y se están cargando correctamente los datos en la dimensión asignada.

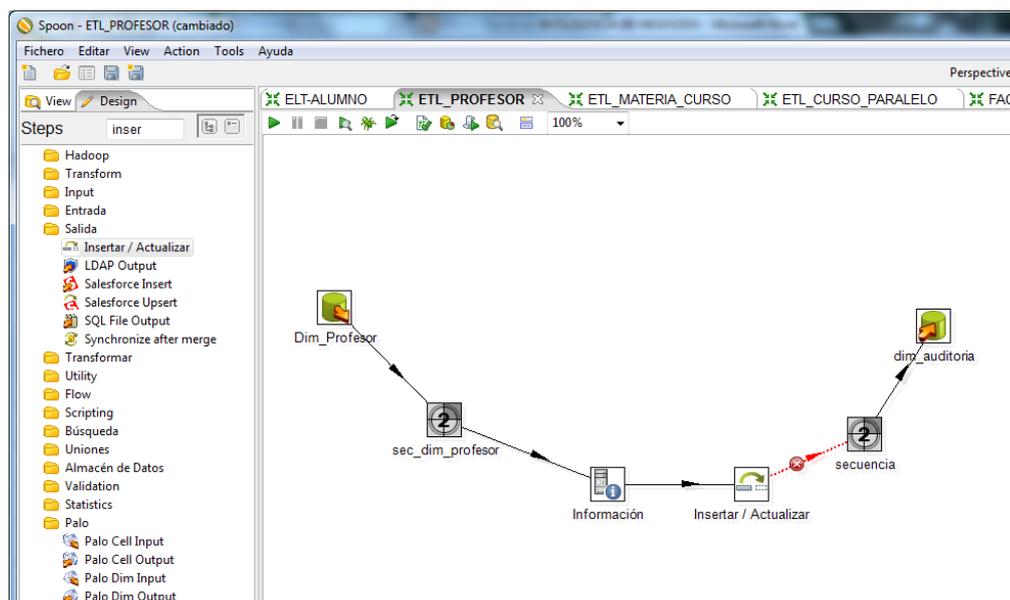


Figura 4.48 Finalización de configuración de una Dimensión completa
Fuente: Elaboración propia

Al finalizar la configuración de cada una de las dimensiones que conformarán el cubo, se proseguirá a configurar la **FACT** que es la tabla principal que está compuesta de todos los **ID'S** de cada una de las tablas dimensionales, a continuación se detallará las herramientas utilizadas:

Table Input.- En ésta herramienta a través de la ejecución de una sentencia SQL se llama a los ID'S de cada una de las tablas de la base de datos transaccional, estos datos serán los que proporcionarán información al cubo.

Al dar doble clic sobre el ícono Tabla de Entrada se desplegará una ventana donde se harán las siguientes configuraciones:

- Nombre de la Tabla de Entrada al cual le denominaremos fact_notas y fact_carrera.
- Conexión con la base de datos.
- Sentencia SQL para extraer los datos.

Configuraciones que ya fueron descritas anteriormente al igual que la conexión con la base de datos y los parámetros, a continuación se mostrarán los datos cargados previamente.

#	id_personal	id_materia_curso	id_curso_paralelo	id_alumno	id_nota_detalle	id_calificacion	id_parcial_institucion	nota	Suma	Contador	Promedio	fecha
1	1072	1006	42	18	1208	25292	17	9	86	9	9.555555555555556	2013/10/28
2	1072	1006	42	18	1208	25082	17	10	86	9	9.555555555555556	2013/10/28
3	1072	1006	42	18	1208	23513	17	9	86	9	9.555555555555556	2013/10/28
4	1072	1006	42	18	1208	25147	17	9	86	9	9.555555555555556	2013/10/28
5	1072	1006	42	18	1208	25015	17	10	86	9	9.555555555555556	2013/10/28
6	1072	1006	42	18	1208	23539	17	10	86	9	9.555555555555556	2013/10/28
7	1072	1006	42	18	1208	24958	17	9	86	9	9.555555555555556	2013/10/28
8	1072	1006	42	18	1208	23585	17	10	86	9	9.555555555555556	2013/10/28
9	1072	1006	42	18	1208	24587	17	10	86	9	9.555555555555556	2013/10/28
10	1072	1007	42	18	1258	8419	17	10	118	12	9.833333333333333	2013/10/23
11	1072	1007	42	18	1258	25434	17	10	118	12	9.833333333333333	2013/10/28
12	1072	1007	42	18	1258	8378	17	10	118	12	9.833333333333333	2013/10/28
13	1072	1007	42	18	1258	2702	17	10	118	12	9.833333333333333	2013/10/17
14	1072	1007	42	18	1258	8536	17	10	118	12	9.833333333333333	2013/10/23
15	1072	1007	42	18	1258	8741	17	10	118	12	9.833333333333333	2013/10/23
16	1072	1007	42	18	1258	8684	17	9	118	12	9.833333333333333	2013/10/28
17	1072	1007	42	18	1258	8587	17	10	118	12	9.833333333333333	2013/10/23
18	1072	1007	42	18	1258	25573	17	10	118	12	9.833333333333333	2013/10/28
19	1072	1007	42	18	1258	25518	17	9	118	12	9.833333333333333	2013/10/28
20	1072	1007	42	18	1258	8495	17	10	118	12	9.833333333333333	2013/10/23
21	1072	1007	42	18	1258	8444	17	10	118	12	9.833333333333333	2013/10/28
22	1071	1009	42	18	899	4965	17	10	49	5	9.8	2013/10/21
23	1071	1009	42	18	899	806	17	10	49	5	9.8	2013/10/16
24	1071	1009	42	18	899	4941	17	9,5	49	5	9.8	2013/10/21
25	1071	1009	42	18	899	22516	17	9,5	49	5	9.8	2013/10/28
26	1071	1009	42	18	899	4989	17	10	49	5	9.8	2013/10/21
27	1072	1011	42	18	1233	20275	17	10	169	17	9.9411764705882353	2013/10/27
28	1072	1011	42	18	1233	20409	17	10	169	17	9.9411764705882353	2013/10/28
29	1072	1011	42	18	1233	17395	17	10	169	17	9.9411764705882353	2013/10/25
30	1072	1011	42	18	1233	17285	17	10	169	17	9.9411764705882353	2013/10/25
31	1072	1011	42	18	1233	17330	17	10	169	17	9.9411764705882353	2013/10/25

Figura 4.49 Visualización de datos cargados en la FACT
Fuente: Elaboración propia

4.3.5.6 SEARCH IN DATABASE O BÚSQUEDA EN BASE DE DATOS.- Ésta herramienta permite buscar valores en la base de datos dimensional.

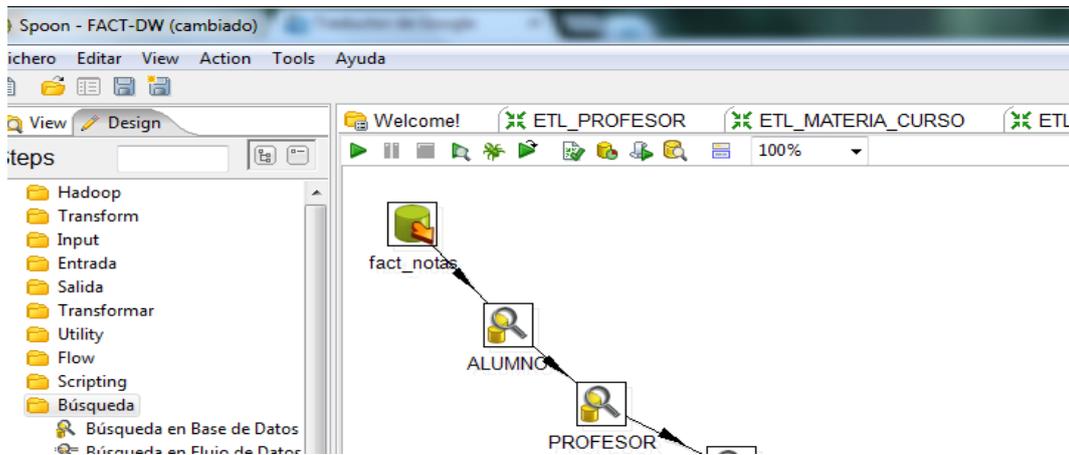


Figura 4.50 Insertando un Search in Database
Fuente: Elaboración propia

Al dar doble clic hay que configurar 3 secciones: la parte de la conexión con la base de datos, la parte de claves para la búsqueda de valores y la parte de valores a devolver de la tabla de búsqueda, a continuación se describirá los siguientes campos:

- **Nombre de paso:** identifica a cada una de las dimensiones creadas.
- **Conexión:** configura la conexión con la base de datos dimensional, este paso ya fue indicado anteriormente.
- **Esquema de búsqueda:** esquema donde está creada la base de datos dimensional.
- **Tabla de búsqueda:** nombre de la tabla dimensional que se encuentra creada en la base de datos dimensional.

Figura 4.51 Configuración de un Search in Database
Fuente: Elaboración propia

Para la segunda parte de la configuración comparar los campos de los ID's.

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

#	Campo de tabla	Comparador	Campo1	Campo2
1	alm_id_alumno	=	id_alumno	

Figura 4.52 Configuración de claves en un Search in Database
Fuente: Elaboración propia

Para la tercera parte de la configuración seleccionar el campo sk de la dimensión y escoger el tipo de dato que va a recibir.

Valores a devolver de la tabla de búsqueda :

#	Campo	Nuevo nombre	Defecto	Tipo
1	alm_sk			Integer

No procesar la fila si la búsqueda falla

¿Producir error si se obtienen múltiples

Ordenar por

Vale Cancelar Obtener Campos Obtener Campos Búsqueda

Figura 4.53 Configuración de valores a devolver en un Search in Database
Fuente: Elaboración propia

Una vez terminada la configuración, este proceso se lo realiza para cada una de las dimensiones que se tengan creadas.

4.3.5.7 MEMORY GROUP BY O GRUPO DE MEMORIA.- Esta herramienta permite agregar campos calculados, no requiere de una entrada clasificada.

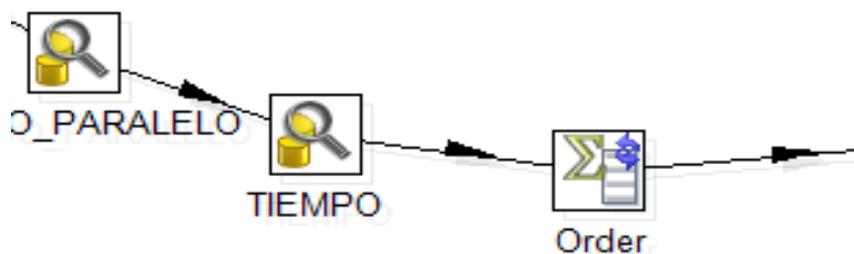


Figura 4.54 Insertando un Memory Group By
Fuente: Elaboración propia

Al dar doble clic en ésta herramienta se configura los siguientes campos:

- **Nombre de paso:** que sirve para identificar el paso a ejecutarse.
- **Campos que forman la agrupación:** aquí se seleccionan todos los campos que contienen los SK de las tablas dimensionales.
- **Agregados:** en esta parte se selecciona el campo a calcularse.

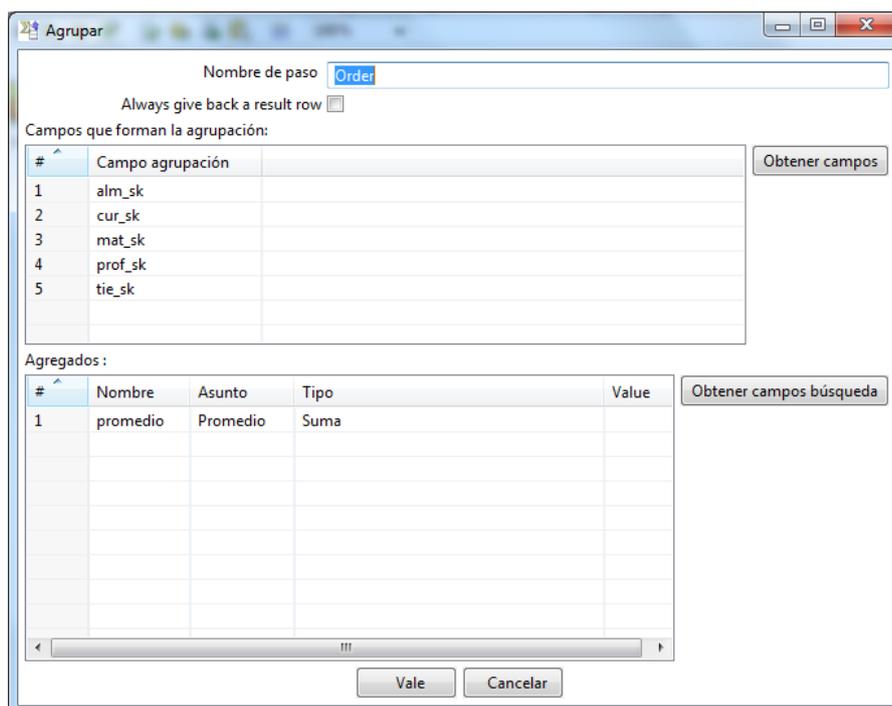


Figura 4.55 Configuración de campos en un Memory Group By
Fuente: Elaboración propia

Luego de realizadas estas configuraciones se agregará un **GET SYSTEM INFO** para obtener los datos propios del sistema, mismos que serán utilizados en la tabla de auditoría, también hay que configurar los campos obligatorios que exige la herramienta, una vez terminado este proceso se agregará un **INSERT/UPDATE** para controlar de manera ordenada la inserción o actualización de la información que va a ser ingresada en la base de datos dimensional también necesita que los campos sean configurados, cabe mencionar que esta herramienta consta de tres bloques de configuración: la parte de conexión con la base de datos, la parte de configuración de claves y la parte de configuración de campos de actualización pasos que ya se describieron anteriormente.

Luego de realizados estos pasos se debe agregar una nueva herramienta de secuencia de errores para finalizar el proceso de **ETL** de cada una de las dimensiones propuestas que conformarán el Datamart, al igual que las demás herramientas necesita ser configurada, cabe mencionar que estos pasos ya fueron descritos anteriormente y finalmente se debe agregar un **TABLE OUTPUT** en la que se almacenaran los errores y los campos de auditoría que se definieron anteriormente, al igual que las demás herramientas se debe de configurar para su funcionamiento, finalmente mostramos la parte de la **FACT** funcionando .

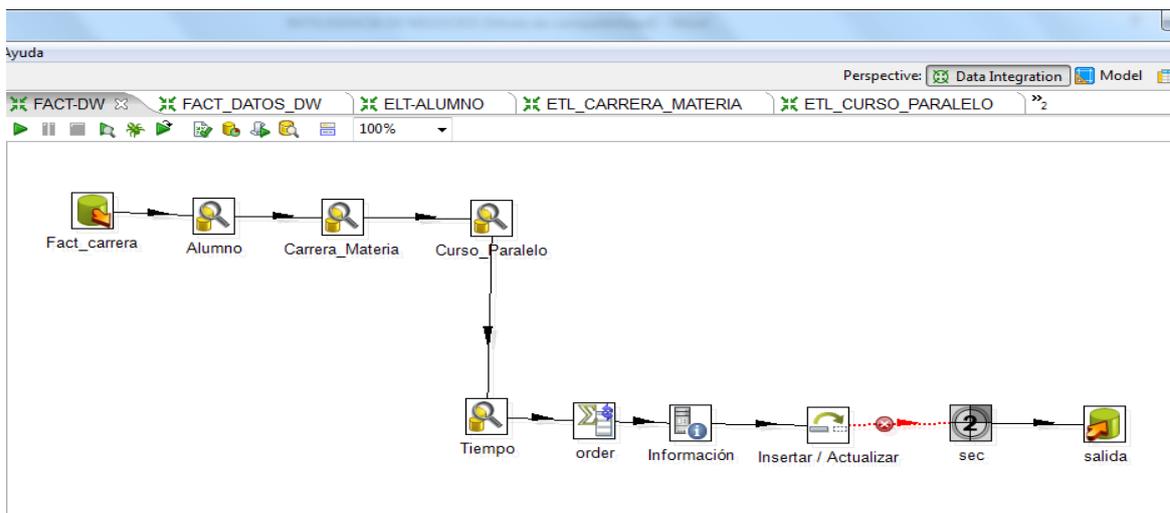


Figura 4.56 Configuración de la **FACT_DW**
Fuente: Elaboración propia

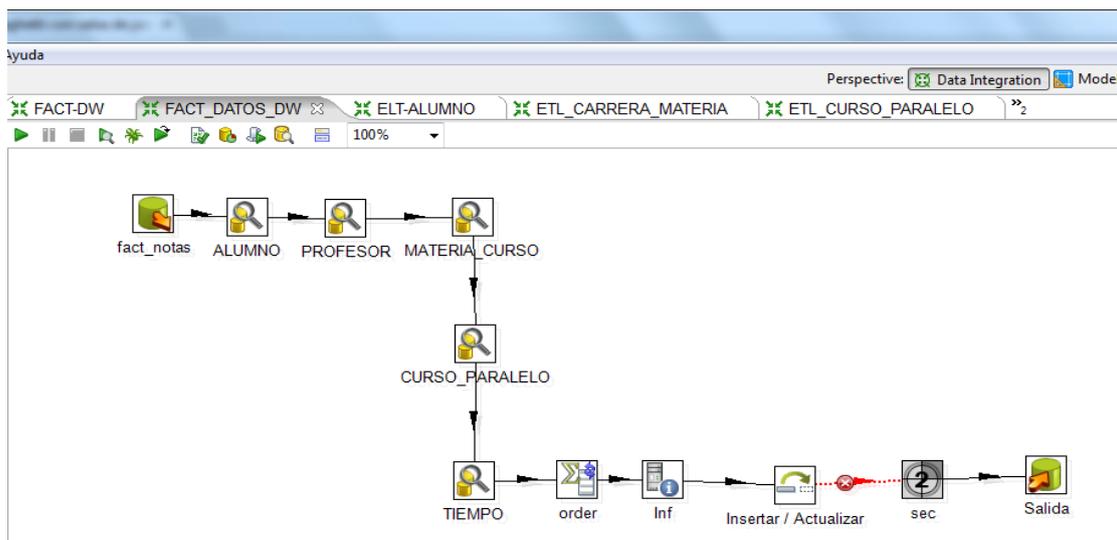


Figura 4.57 Configuración de la **FACT_DATOS_DW**
Fuente: Elaboración propia

4.3.5.8 Diseño del Cubo de Información

Para el diseño del cubo multidimensional se utilizará la herramienta Schema Workbench es necesario tener el proceso ETL listo y el Datamart debe contener información, se debe ingresar a la herramienta y a continuación deberá agregar un nuevo cubo:

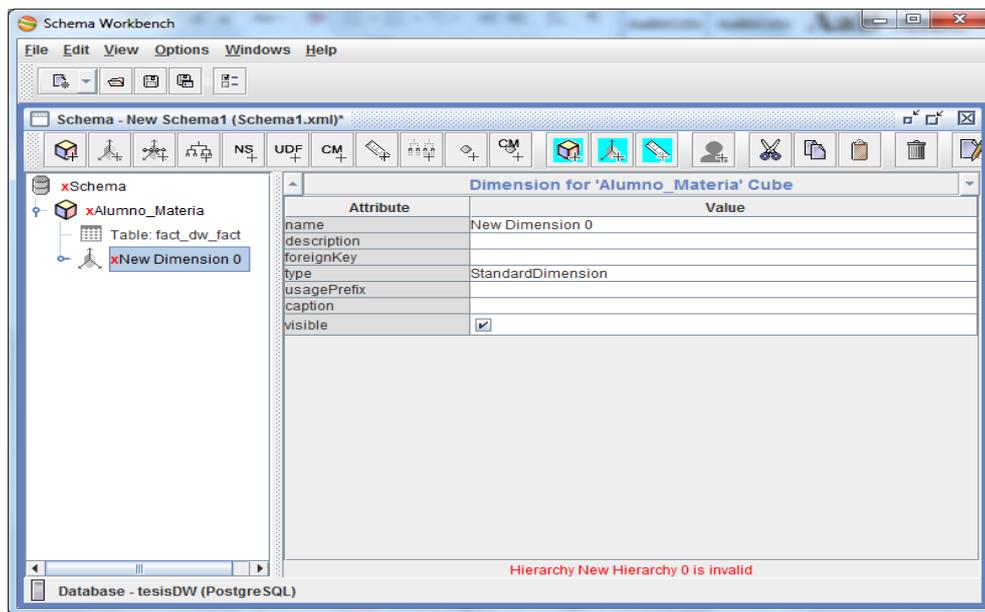


Figura 4.58 Creación del nuevo Cubo Alumno_Materia
Fuente: Elaboración propia

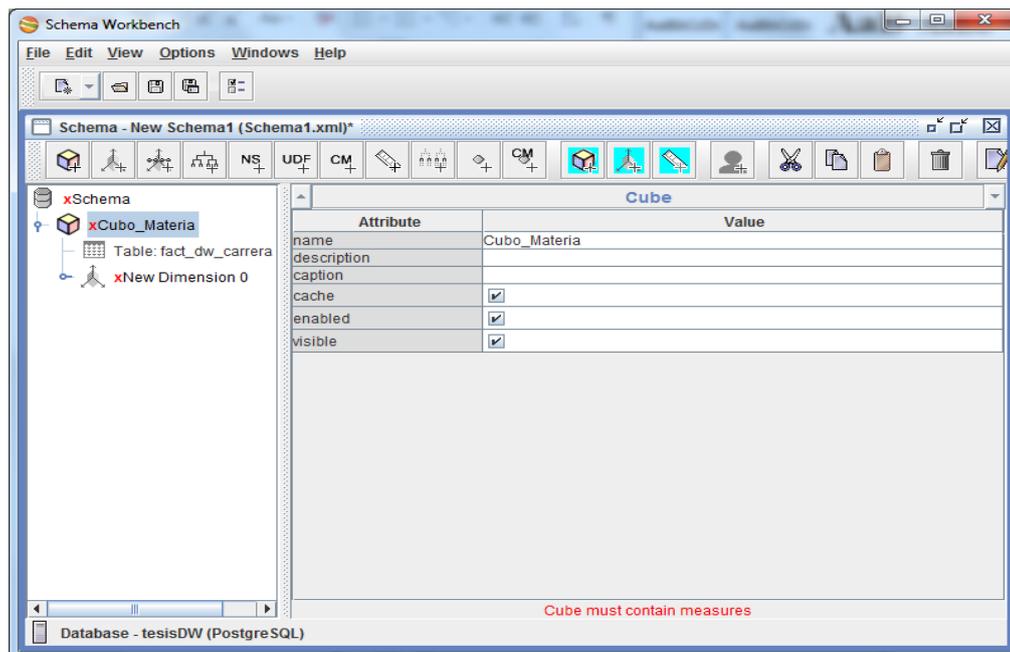


Figura 4.59 Creación del nuevo Cubo Cubo_Materia
Fuente: Elaboración propia

Una vez creado el cubo se deberá agregar una nueva dimensión, es necesario tomar en cuenta que aquí deberá agregar todas las dimensiones que creó en el proceso ETL, en la siguiente imagen se muestra como se añade la tabla de hechos y cada una de las dimensiones con sus respectivos niveles, es decir se creó un nivel de jerarquía para cada una de las dimensiones y se agregó el campo calculado.

Aquí se deberá definir el orden en el que aparecerán las dimensiones al momento de publicar el cubo, también es importante configurar cada uno de los campos tanto en las jerarquías como en las dimensiones para que no surjan errores al momento de publicar los datos.

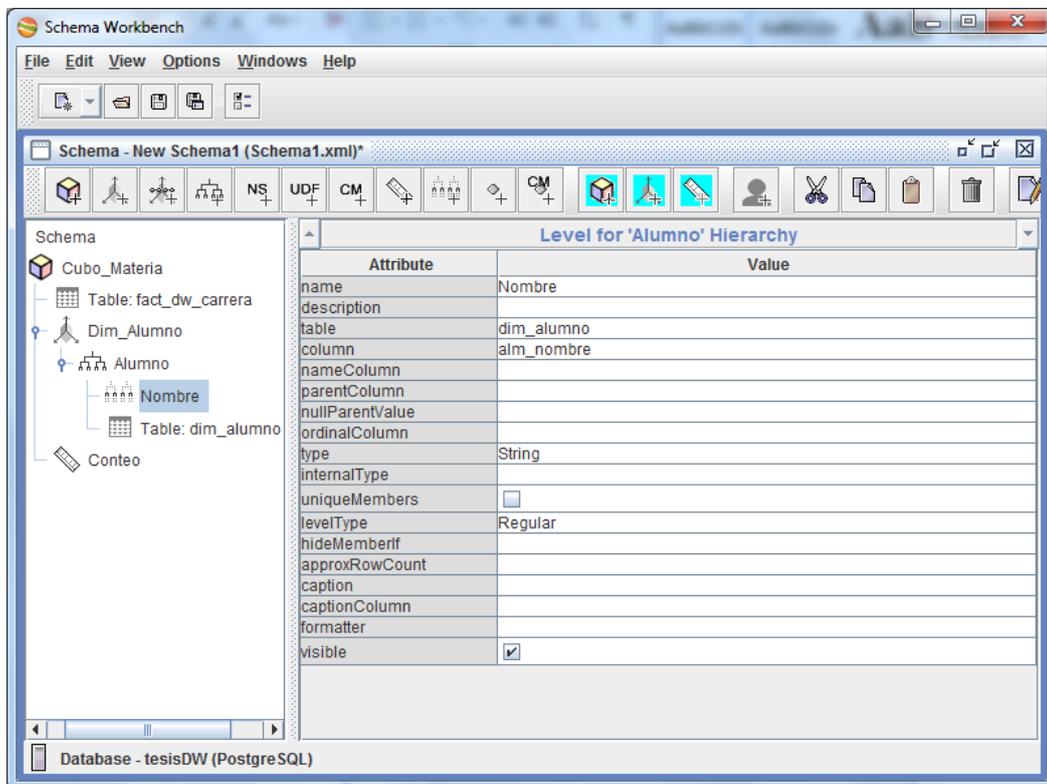


Figura 4.60 Creación de dimensiones y jerarquías en el nuevo Cubo
Fuente: Elaboración propia

A continuación se presenta a cada uno de los cubos diseñados:

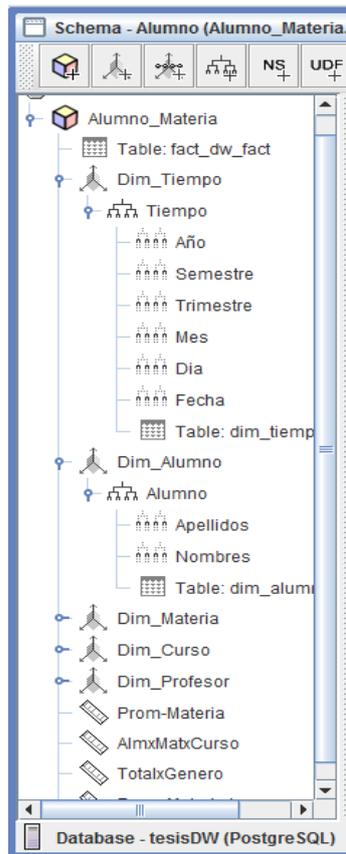


Figura 4.61 Cubo completo Alumno_Materia
Fuente: Elaboración propia

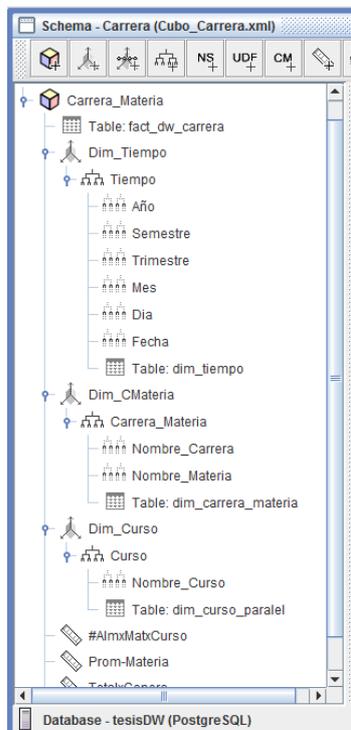


Figura 4.62 Cubo completo Cubo_Carrera
Fuente: Elaboración propia

Luego de haber terminado todo este proceso se procederá a publicar el cubo de información, para ello es necesario dar clic en **FILE** y deberá escoger la opción **PUBLISH**, a continuación se desplegará una ventana donde deberá especificar el usuario y la contraseña mismos que deberán ser creados en el servidor Mondrian:

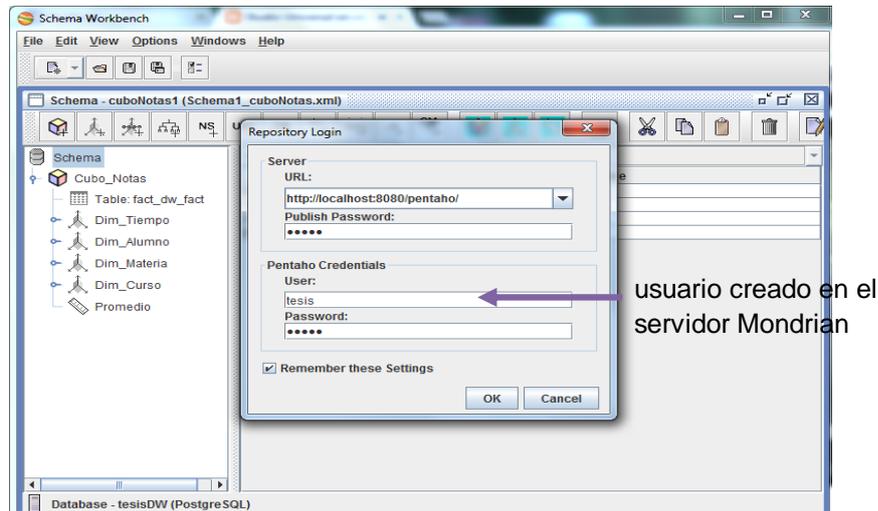


Figura 4.63 Configuración de usuario para publicación del Cubo
Fuente: Elaboración propia

Luego de dar clic en **OK**, se desplegará otra ventana donde se deberá escoger la ubicación del directorio donde se publicará el cubo.

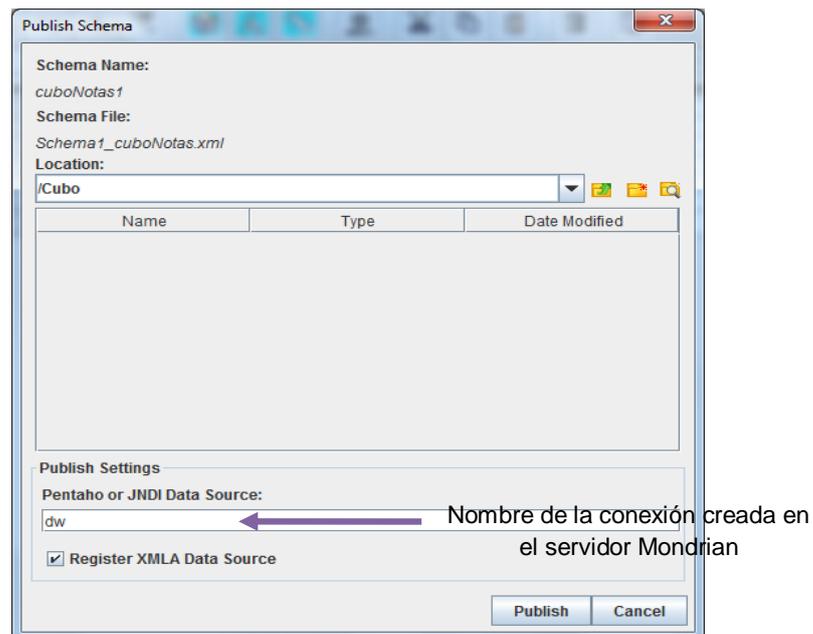


Figura 4.64 Selección de ubicación para publicación del Cubo
Fuente: Elaboración propia

Una vez finalizado el proceso de selección de directorio se procederá a dar clic en **PUBLISH**, si está correcta la configuración el cubo se publicará sin errores.

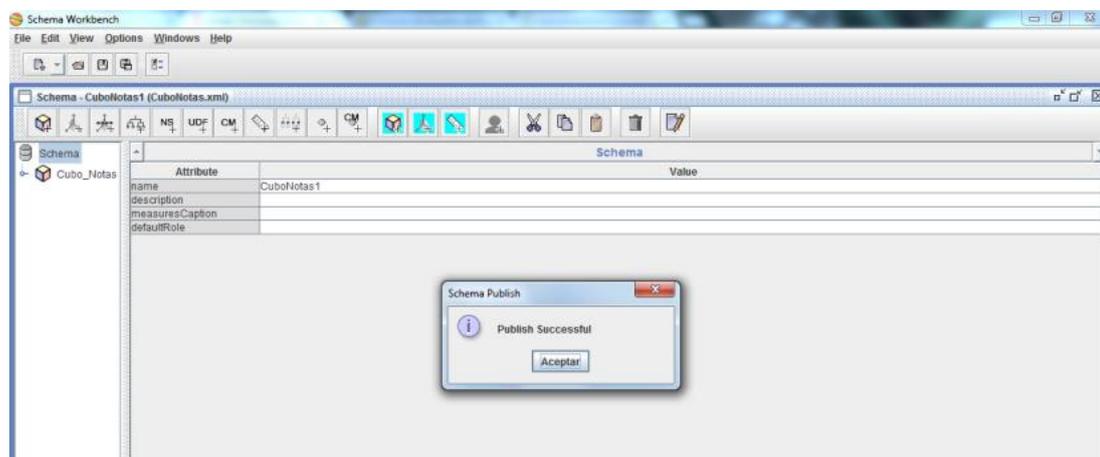


Figura 4.65 Publicación del Cubo
Fuente: Elaboración propia

4.3.5.9 Publicación y visualización del cubo del información

Para la publicación de los cubos de información es necesario utilizar la herramienta Biserver que es parte de la suite de Pentaho, esta herramienta está compuesta de una carpeta denominada Biserver-ce que contiene el framework del Servidor y la carpeta Administration-console que contiene los archivos de inicio para la administración de la consola, ambas carpetas ayudan a levantar los servicios del Servidor Mondrian, esta herramienta permite la creación y asignación de usuarios y roles. El servidor de Pentaho está compuesto de librerías y una plataforma de inteligencia de negocios, éste servidor se ejecuta en un servidor web J2EE como el Apache Tomcat, JBOSS AS, WebShpere, WebLogic y Oracle As y contiene motores para realizar reportes, análisis de información, minería de datos, etc.

Para ingresar a la consola de administración, deberá abrir una ventana de un navegador web y por medio de la dirección IP con la que esté trabajando y el puerto por defecto que utiliza la consola de administración por ejemplo: ip_maquina:8099, ingresará a una página donde le permitirá realizar una serie de actividades entre las más importantes está la creación de usuarios, asignación de permisos y roles a cada uno de los usuarios creados, conexión con la base de datos, administración de grupos, etc.

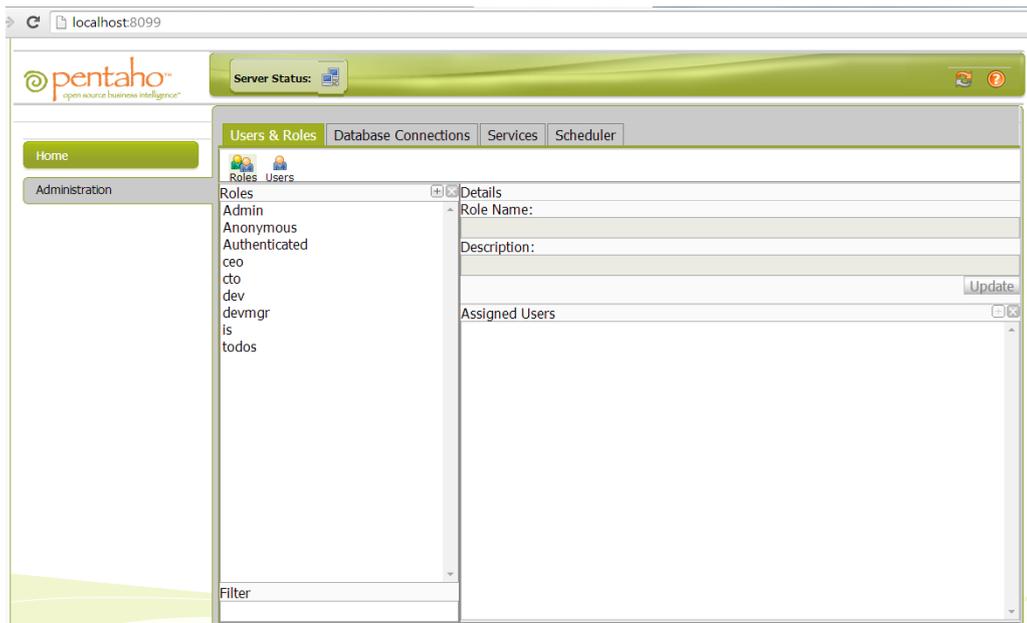


Figura 4.66 Página principal de la consola de administración
Fuente: Elaboración propia

Para la configuración de la conexión con la base de datos deberá escoger la pestaña **DATABASE CONNECTIONS**, allí se desplegará una ventana en la cual deberá ingresar todos los datos que exige.

Figura 4.67 Configuración con la base de datos
Fuente: Elaboración propia

Para ingresar al servidor se lo hará a través de una ventana de un navegador web por medio de la dirección IP con la que esté trabajando y el puerto por defecto que utiliza el servidor por ejemplo: ip_maquina:8080/pentaho, presenta una pantalla de login donde ingresará el usuario creado en la consola de administración.



Figura 4.68 Página inicial del Servidor Mondrian
Fuente: Elaboración propia

Al ingresar muestra un menú principal de la herramienta Pentaho Bi Server, permite escoger las opciones de nuevos reportes, análisis, visualización de ejemplos y visualización de cubos generados y publicados por el usuario para su respectivo análisis a través de la herramienta Schema Work Bench.

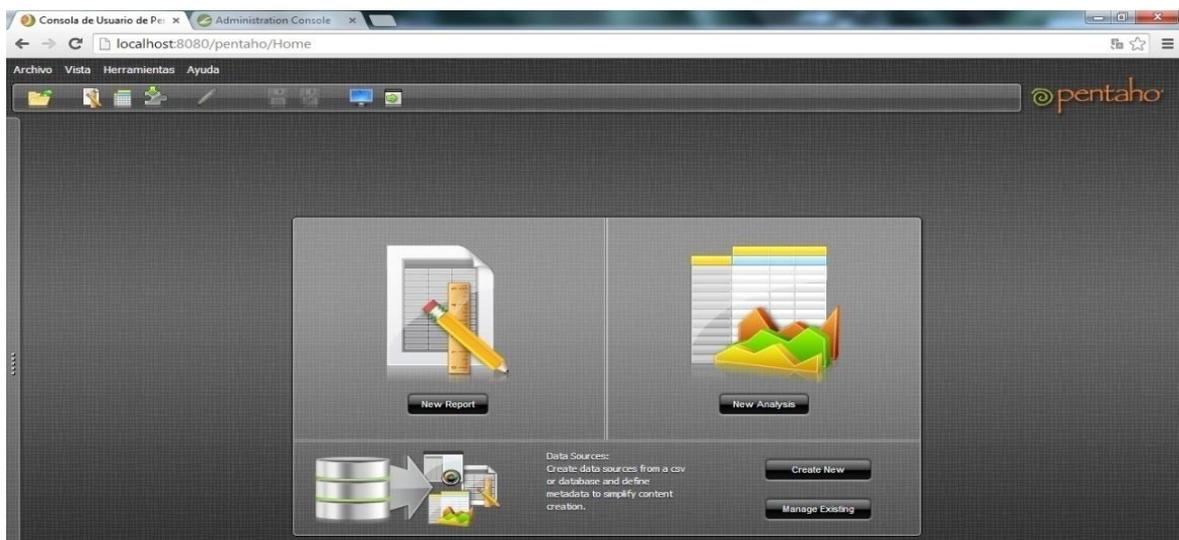


Figura 4.69 Menú principal de Pentaho Bi Server
Fuente: Elaboración propia

La herramienta permite realizar el análisis de diferentes cubos para ellos es necesario seleccionar el botón **NEW ANALYSIS VIEW**, aquí deberá escoger el nombre del esquema donde está creado el cubo y seleccionar el nombre del cubo creado en el caso de existir varios cubos en un mismo esquema.

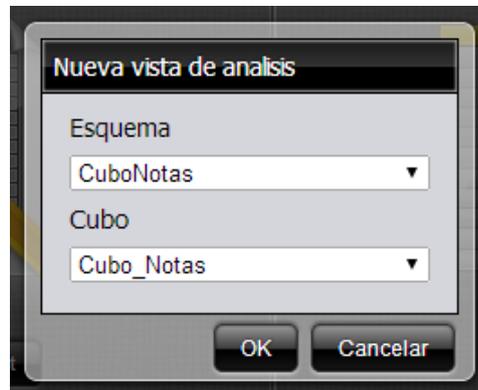


Figura 4.70 Creación de una nueva vista de análisis
Fuente: Elaboración propia

Automáticamente el cubo de información se mostrará con su información completa y el usuario podrá elaborar vistas de análisis, reportes, exportar la información a Excel y guardar esta información en una carpeta personalizada.

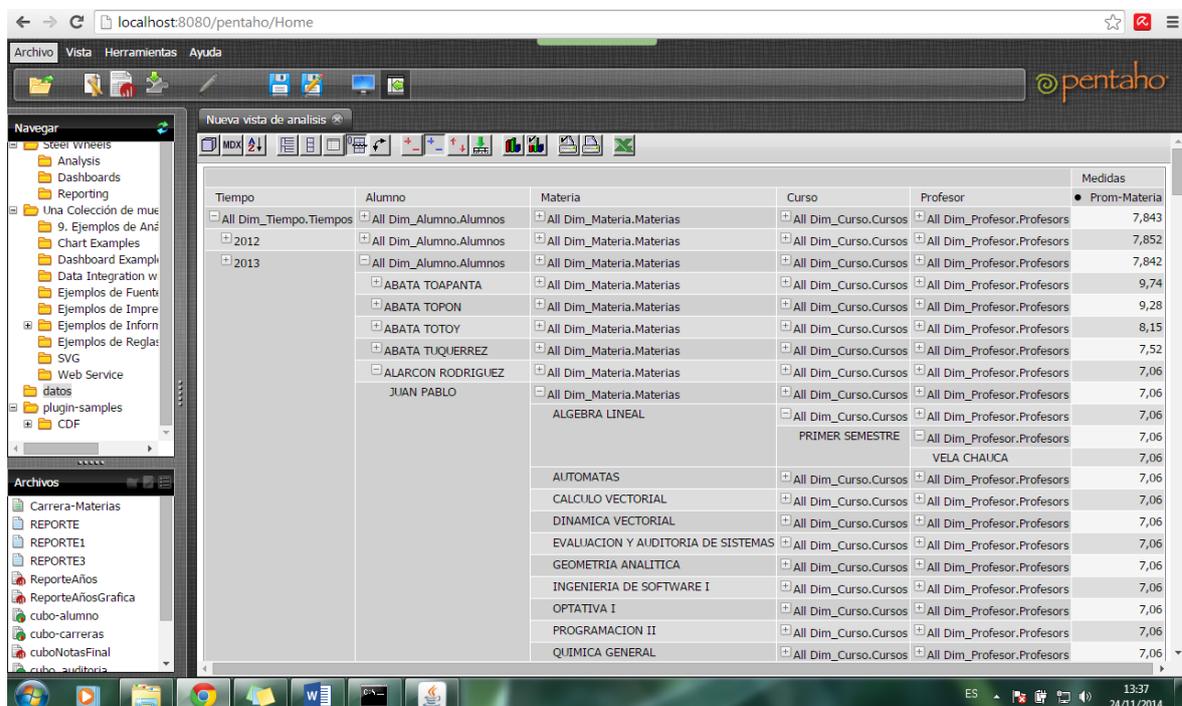


Figura 4.71 Vista del cubo de información Alumno_Materia
Fuente: Elaboración propia

Medidas	Curso	Carrera_Materia	Tiempo
#AlmMabCurso	Prom.Materia1	TotalxGenero	Prc
33.388	8,87	22.992.504	
3.144	8,911	4.977.216	2012
255	8,825	402.840	Carrera de Ingeniería Textil
560	8,288	884.280	Carrera de Ingeniería en Electronica y Redes de Comunicación
871	9,023	1.379.376	Carrera de Ingeniería en Mecatronica
1.458	9,098	2.310.720	Carrera de Ingeniería en Sistemas Computacionales
30	9,12	47.688	APLICACIONES INFORMATICAS I
30	9,12	47.688	SEPTIMO SEMESTRE
136	9,66	215.904	ARQUITECTURA DE SOFTWARE
222	8,88	352.416	AUTOMATAS
43	9,25	67.992	COMPILADORES
119	9,63	187.992	COSTOS
39	9,62	61.176	DERECHO INFORMatico
179	9,1	284.424	EVALUACION Y AUDITORIA DE SISTEMAS
23	9,28	35.496	GERENCIA DE EMPRESAS
139	9,3	219.696	INGENIERIA DE SOFTWARE I
53	8,44	83.184	INGENIERIA ECONOMICA
16	8,61	25.680	PRODUCCION DE AUDIO Y VIDEO
238	9,03	378.096	PROGRAMACION II
15	8,21	23.976	PROGRAMACION VI
206	8,71	327.000	REDES I

Figura 4.72 Vista del cubo de información Carrera_Materia
Fuente: Elaboración propia

Para facilitar el análisis de la información se hizo la instalación de un plugin en la consola de administración, se llama **Saiku** éste se encuentra disponible en Pentaho Marketplace, se hace la descarga y automáticamente aparece en la consola de administración como podemos observar en la siguiente figura:

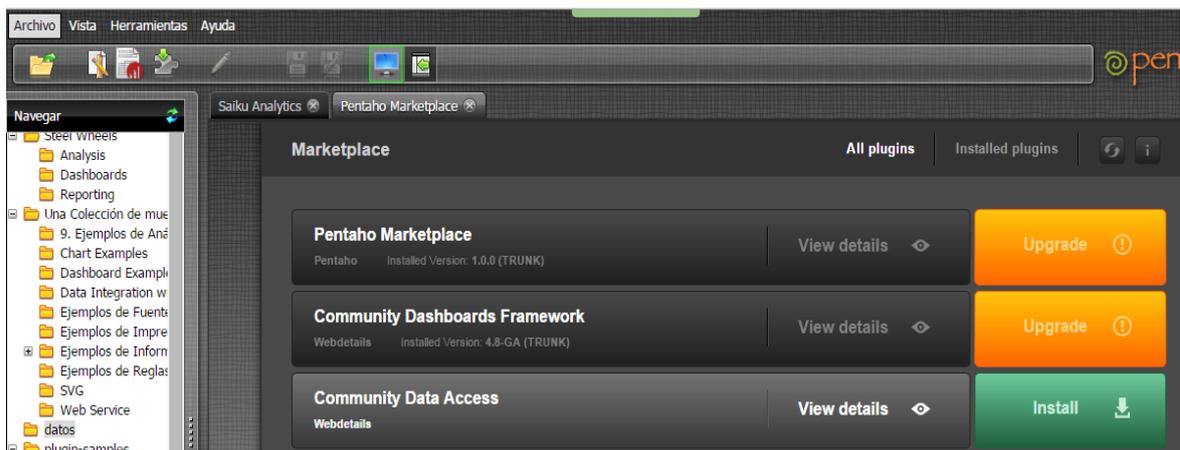


Figura 4.73 Instalación del plugin Saiku en la consola de administración
Fuente: Elaboración propia

Saiku permite a los usuarios explorar fuentes de datos complejas, dentro de un navegador, una vez hecho el análisis se puede guardar los resultados, exportarlos a Excel o PDF, todo directamente desde el navegador, esta herramienta se distribuye bajo la licencia Apache 2.0, que es licencia de código abierto empresarial, la funcionalidad que trae ésta herramienta es que se pueden sacar totales, la desviación estándar, la media aritmética, realizar filtrados de datos, seleccionar jerarquías, además de una interfaz amigable. **(Meteorite Consulting, 2014)**

Saiku se conecta a una amplia gama de fuentes de datos que le permite explorar los datos en tiempo real directamente desde la fuente, se puede integrar fácilmente en los marcos de seguridad y está optimizado para funcionar con el hardware del servidor de los productos básicos, el almacenamiento en caché inteligente reduce el impacto en el rendimiento de la base de datos subyacente y minimiza el tráfico de red. Algunos ejemplos de fuentes de datos podemos mencionar los siguientes: Microsoft SQL Server, Microsoft Analysis Services, base de datos Oracle, Oracle Essbase, MongoDB, MySQL, PostgreSQL, Cloudera Impala, Vectorwise Actian, Amazon Redshift, Teradata, Vertica, etc.

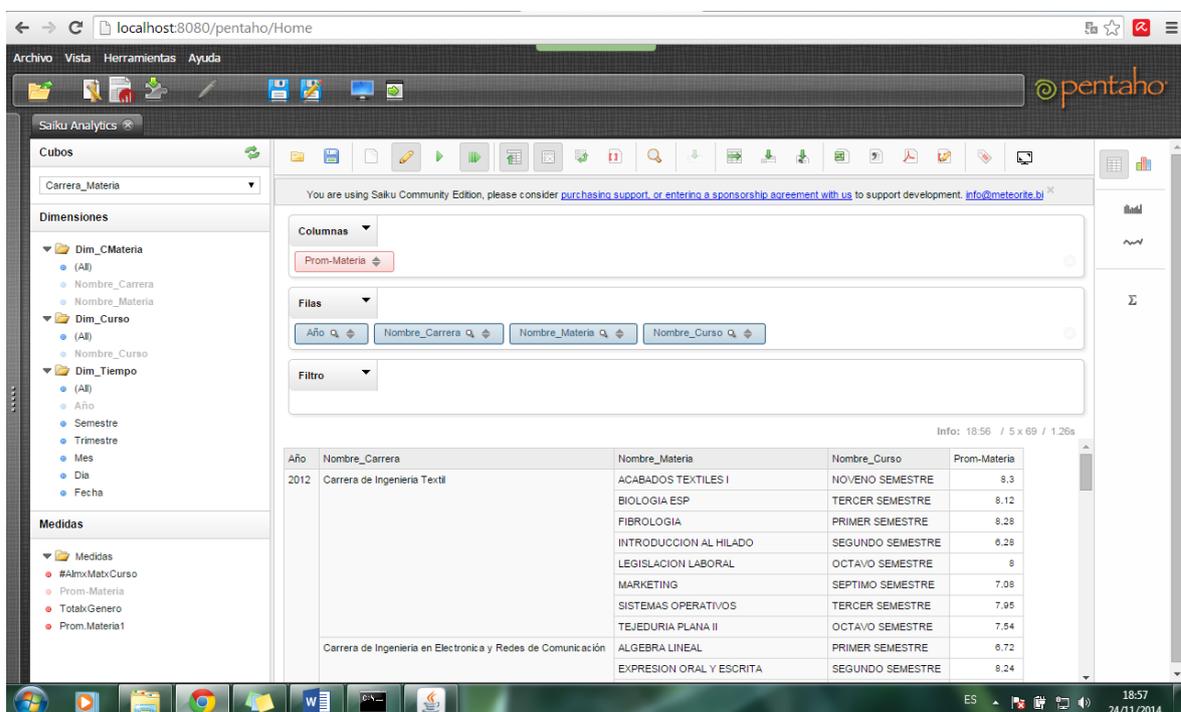


Figura 4.74 Vista del cubo con Saiku en la consola de administración
Fuente: Elaboración propia

4.3.5.10 Generación de reportes

Para la generación de reportes se utilizará la herramienta Pentaho Reporting Services que es capaz de generar informes programáticos sobre la base de un archivo de definición XML, sobre esta solución se han desarrollado muchas herramientas, por ejemplo informes, diseñadores de interfaz gráfica de usuario, y asistentes tipo wizard.

Pentaho Reporting es un conjunto de herramientas de informes de código abierto que permite crear informes relacionales y análisis de una amplia gama de fuentes de datos, es capaz de crear archivos PDF, Excel, HTML, texto, texto enriquecido, XML y CSV. El motor de informes es una biblioteca de informes Java de peso ligero con una huella global mínima, lo que puede ser incrustado en otras aplicaciones Java de escritorio o servidor, para la publicación de los reportes e informes lo hace a través de internet. **(Thomas, 2014)**

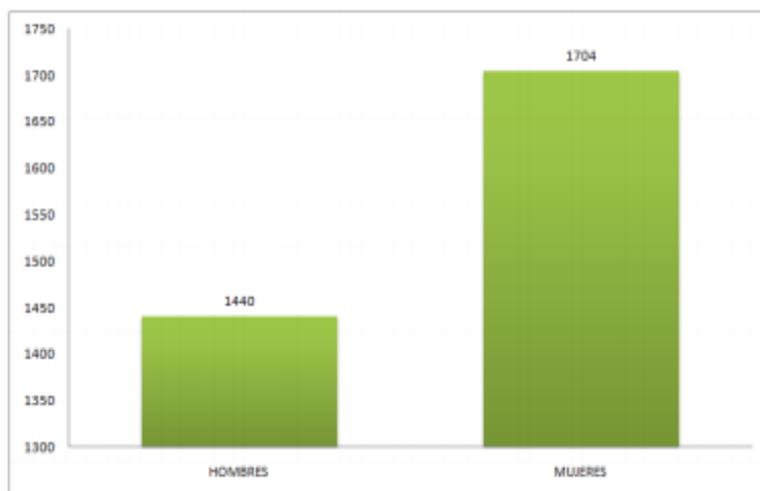
La cantidad de memoria necesaria para el procesamiento de informes depende en gran medida del tamaño del informe y el formato de exportación deseado, es decir, para los informes de tamaño razonable de no más de 2.000 páginas, 512 MB debería ser más que suficiente y se debería añadir 128 MB por cada usuario concurrente adicional. Si hay la necesidad de exportar los informes resultantes de documentos aún más grandes, es necesario añadir memoria adicional o sintonizar el informe de configuración para activar funciones adicionales para reducir el uso de la memoria, a expensas de la velocidad de procesamiento. **(Thomas, 2014)**

Los requisitos mínimos son:

- Java 5 o 6
- Una CPU rápida y actual
- Mínimo 512MB asignada para una sola aplicación

A continuación se muestran algunos informes generados a través de Pentaho Reporting Services y publicados en la consola de administración del Biserver.

ANALISIS ALUMNOS



TOTAL ALUMNOS HOMBRES: 1440

TOTAL ALUMNOS MUJERES: 1704

TOTAL ALUMNOS 3144

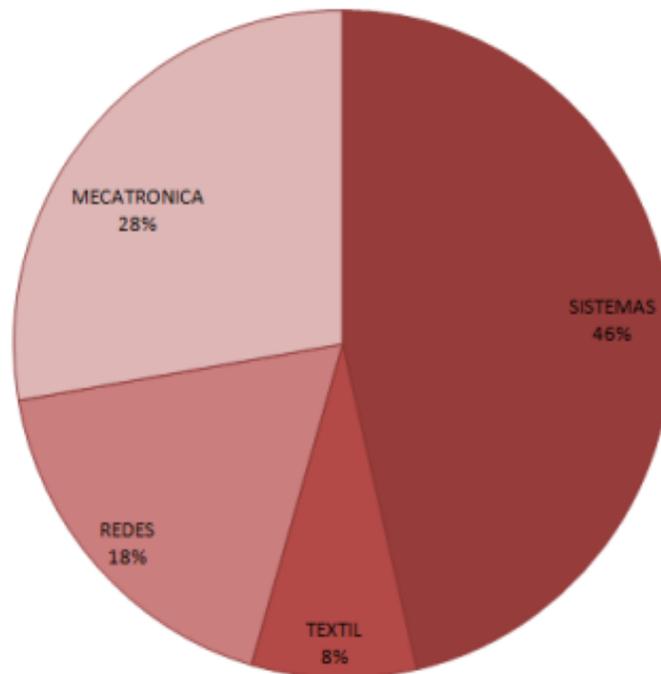
Figura 4.75 Vista del informe generado Total alumnos
Fuente: Elaboración propia

REPORTE DE ALUMNOS CON RELACION A LA MATERIA

TOTAL CARRERAS: 5

TOTAL ESTUDIANTES POR CARRERA: 3144

Sistemas:	1458
Textil:	255
Redes:	560
Mecatronica:	871



PROMEDIO POR MATERIA POR CURSO: 7.8

Figura 4.76 Vista del informe generado alumnos con relación a la materia
Fuente: Elaboración propia

CARRERAS CON SUS RESPECTIVAS MATERIAS

cmat_nombre_carrera: Carrera de Ingeniería en Electrónica y Redes de Comunicación

cmat_nombre_materia	cmat_num_creditos
ADMINISTRACION Y GESTION REDES	4
ALGEBRA LINEAL	6
ANALISIS DE SEÑALES	4
ANALISIS MATEMATICO II	6
BASE DE DATOS	4
CABLEADO ESTRUCTURADO	4
CIRCUITOS ELECTRONICOS	4
COMUNICACION ANALOGA	4
COMUNICACION DIGITAL	6
COMUNICACION INHALAMBERICA	4
COMUNICACION SATELITAL	4
DERECHO APLICACION TELECOMUNICACIONES	4
DISEÑO SISTEMAS DE COMUNICACIONES	4
ECUACIONES DIFERENCIALES	6
ELECTRONICA	6
ENERGIAS ALTERNATIVAS	4
ESTATICA	6
EXPRESION ORAL Y ESCRITA	4
GEOMETRIA	4
GEOMETRIA ANALITICA	6
INSTALACIONES ELECTRICAS	4
MATEMATICA APLICADA	4
MICROS AVANZADOS	4
NETWORKING I	6
NETWORKING II	6
NETWORKING III	6
NGN	4
PROB. Y PROCESOS ESTADISTICOS	4
PROG. DE SISTEMAS MULTIMEDIA	5
PROYECTOS DE REDES	4
ROBOTICA E INTELIGENCIA ARTIFICIAL	4
SEGURIDAD DE REDES	4
SIMULACION DE PROCESADORES	4
SISTEMA DE FIBRA OPTICA	4
SISTEMAS ANALOGOS DIGITALES	4
SISTEMAS DE COMUNICACION RADIANTE	5
SISTEMAS DE TRANSMISION	4
SISTEMAS DIGITALES	6
TEORIA ELECTROMAGNETICA	5
TRATAMIENTO DIGITAL DE SEÑALES	4
WLAN	4
41 Total de materias	Total de créditos 189

cmat_nombre_carrera: Carrera de Ingeniería en Mecatrónica

cmat_nombre_materia	cmat_num_creditos
ANALISIS MATEMATICO II	4
APLICACIONES ELECTRONICAS	6
APLICACIONES ELECTRONICAS	6
AUTOMATIZACION INDUSTRIAL	6
CALCULO VECTORIAL	10
CIENCIA DE LOS MATERIALES	6
CIRCUITOS ELECTRICOS	6
CIRCUITOS ELECTRICOS II	6
DIBUJO MECANICO I	6
DIBUJO MECANICO II	6
DINAMICA VECTORIAL	6
DISEÑO MECANICO	6
DISEÑO MECANICO I	6
DISEÑO MECATRONICO I	6
DISEÑO MECATRONICO II	6

ELECTRONICA DE POTENCIA	6
EMPRENDIMIENTO E INNOVACION TECN.	4
GEOMETRIA PLANA	6
INGENIERIA DE MANUFACTURA	6
INSTRUMENTACION	6
INTRODUCCION A LA INGENIERIA	6
LIDERAZGO Y MOTIVACION	6
MANTENIMIENTO INDUSTRIAL	6
MAQUINAS HERRAMIENTAS	6
MECANICA DE FLUIDOS	6
MECANISMOS	6
NEUMATICA	6
OPTATIVA I	6
PROYECTOS DE DESARROLLO	4
QUIMICA GENERAL	6
REDES INDUSTRIALES	6
SISTEMAS DE CONTROL I	6
SISTEMAS DE CONTROL II	4
SISTEMAS LINEALES	6
SISTEMAS MICROPROCESADOS	6

35 Total de materias Total de créditos **206**

cmat_nombre_carrera: Carrera de Ingenieria en Sistemas Computacionales

cmat_nombre_materia	cmat_num_creditos
ADMINISTRACION DE EMPRESAS	6
ADMINISTRACION DE SISTEMAS I	4
ADMINISTRACION DE SISTEMAS II	6
ANALISIS MATEMATICO I	6
ANALISIS NUMERICO	6
APLICACIONES INFORMATICAS I	6
APLICACIONES INFORMATICAS II	6
ARQUITECTURA DE COMPUTADORAS	6
ARQUITECTURA DE SOFTWARE	6
AUTOMATAS	6
BASE DE DATOS I	6
BASE DE DATOS II	6
BASE DE DATOS III	6
CIRCUITOS DIGITALES	6
COMPILADORES	6
COMPUTACION MOVIL	6
COSTOS	6
DERECHO INFORMATICO	6
ESTRUCTURA DE DATOS II	4
EVALUACION Y AUDITORIA DE SISTEMAS	6
GERENCIA DE EMPRESAS	6
GERENCIA DE EMPRESAS II	6
GRAFICACION Y ANIMACION	6
HIGIENE SEGURIDAD Y AMBIENTE	6
INFORMATICA GENERAL	6
INGENIERIA DE SOFTWARE I	6
INGENIERIA DE SOFTWARE II	4
INGENIERIA ECONOMICA	6
INTELIGENCIA ARTIFICIAL	6
MATEMATICAS DISCRETAS	6
OPTATIVA II	6
OPTATIVA III	6
OPTATIVA IV	6
PROBABILIDAD Y ESTADISTICA	6
PRODUCCION DE AUDIO Y VIDEO	6
PROGRAMACION I	6
PROGRAMACION II	6
PROGRAMACION III	4
PROGRAMACION IV	4
PROGRAMACION V	4
PROGRAMACION VI	4
REALIDAD NACIONAL	4
REDES I	6
REDES II	6

SEGURIDAD INFORMATICA	6
SISTEMAS DE INF. GEOGRAFICA	6
SISTEMAS MULTIMEDIA	6
TECNICAS DE APRENDIZAJE	4
TECNOLOGIA ELECTRICA	6
TRABAJO DE GRADO I	4
TRABAJO DE GRADO II	6

51 Total de materias Total de créditos **286**

cmat_nombre_carrera: Carrera de Ingenieria Industrial

cmat_nombre_materia	cmat_num_creditos
ACUSTICA	4
COMPUTACION	4
CONTABILIDAD DE COSTOS	4
CONTROL AUTOMATICO	4
COSTEO BASADO EN ACTIVIDADES	4
DIRECCION DE MARKETING	4
DISEÑO DE PRODUCTOS Y SERVICIOS	4
ECONOMIA EMPRESARIAL	6
ERGONOMIA Y SEGURIDAD	4
ESTADIS. PARA EL MANEJO DE LA CALID. Y SEIS SIGMA	6
ESTADISTICA APLICADA	4
ESTADISTICA INFERENCIAL	4
GERENCIA DE CALIDAD Y PRODUCTIVIDAD	6
GERENCIA DE OPERACIONES	6
INGENIERIA DE LA PRODUCCION I	5
INGENIERIA DE LA PRODUCCION II	5
INGENIERIA DE PROCESOS	4
INGENIERIA DEL TRABAJO	4
INGENIERIA ECONOMICA Y DE COSTOS	4
LEGISLACION Y DERECHO AMBIENTAL	4
LOGISTICA Y DISTRIBUCION	4
MEDICION DEL TRABAJO YV ESTANDARES DE TIEMPO	6
MICROECONOMIA	4
NORMALIZACION Y METODOLOGIA	6
OPT I (PROCESOS INDUSTRIALES)	4
OPT II(MAQUINAS NEUMATICAS E HIDRAULICAS)	4
OPT III(TEORIA DE INVENTARIOS)	4
PLANEAMIENTO Y CONTROL DE LA PRODUCCION	4
PLANIFICACION DE INSTALACIONES	4
PRODUCCION MAS LIMPIA	4
RECURSOS HUMANOS	4
SEGURIDAD DEL TRABAJO I	5
SEGURIDAD DEL TRABAJO II	6
SIMULACION DE SISTEMAS DE MANUFACTURA	5
SISTEMAS DE GESTION DE CALIDAD Y MEDIOAMBIENTE	6
SISTEMAS Y TECNOLOGIAS DE LA INFORMACION	4

36 Total de materias Total de créditos **164**

cmat_nombre_carrera: Carrera de Ingenieria Textil

cmat_nombre_materia	cmat_num_creditos
ACABADOS TEXTILES I	4
ACABADOS TEXTILES II	4
ANALISIS MATEMATICO III	6
BIOLOGIA	4
BIOLOGIA ESP	4
CONTROL AUTOMATICO	4
CONTROL DE PROCESOS I	4
CONTROL DE PROCESOS II	4
DISEÑO TEXTIL	4
DISEÑO, MODA Y CONFECCION	4

ELECTRICIDAD	4
ELECTRONICA GENERAL	4
FIBROLOGIA	4
FISICA I	4
FISICA II	4
HILATURA ACRILICA	4
HILATURA ALGODONERA I	6
HILATURA ALGODONERA II	6
HILATURA LANERA	4
INTROD. TINTORERIA Y ACABADOS	4
INTRODUCCION AL HILADO	4
INTRODUCCION AL TEJIDO	4
INVESTIGACION CIENTIFICA	6
INVESTIGACION DE OPERACIONES	4
LEGISLACION LABORAL	4
MANTENIMIENTO	4
MARKETING	4
NO TEJIDOS	4
NORMALIZACION INDUSTRIAL	4
PROYECTOS	4
QUIMICA INORGANICA I	8
QUIMICA INORGANICA II	4
QUIMICA ORGANICA II	4
QUIMICA TEXTIL	4
RELACIONES INDUSTRIALES	4
SISTEMAS OPERATIVOS	6
TEJEDURIA PLANA I	4
TEJEDURIA PLANA II	4
TEJEDURIA PUNTO I	4
TEJEDURIA PUNTO II	4
TERMODINAMICA	4
TERMODINAMICA APLICADA	6
TINTORERIA I	4
TINTORERIA II	4
TINTORERIA III	4
45 Total de materias	Total de créditos 196
208	1041

Total Num Materias Total Num Creditos

Page 1 / 1

Figura 4.77 Listado de Carreras con su respectivas Materias
Fuente: Elaboración propia

REPORTE DE LISTADO DE CARRERAS

cmat_nombre_carrera
Carrera de Ingeniería en Electronica y Redes de Comunicación
Carrera de Ingeniería en Mecatronica
Carrera de Ingeniería en Sistemas Computacionales
Carrera de Ingeniería Industrial
Carrera de Ingeniería Textil
5

TOTAL DE CARRERAS:

Page 1 / 1

Figura 4.78 Listado de Carreras
Fuente: Elaboración propia

4.4 PUESTA EN MARCHA

En la mayoría de empresas se desarrolla una gran cantidad de proyectos de diferente índole, desde los más sencillos hasta los proyectos que implican la estructura de la organización, el personal encargado de la gestión de proyectos debe hacer uso de las herramientas y establecer procedimientos estandarizados con la finalidad de optimizar tiempo y lograr el éxito de los proyectos; es importante tomar en cuenta que el proyecto no ha finalizado mientras no se haya realizado la puesta en marcha, se hayan realizado las respectivas pruebas y se hayan obtenido los resultados finales del proyecto.

A continuación, se describirán los requisitos que debe cumplir un proyecto durante la puesta en marcha y las actividades que el personal encargado de realizar las pruebas deberá evaluar:

- Deberá llevar a cabo un monitoreo continuo del proyecto, evaluar parámetros como: calidad, seguridad, mantenimiento, eficiencia, tiempo de respuesta de las diferentes aplicaciones.
- Documentar la información necesaria de parte del equipo evaluador al personal encargado del manejo del proyecto, se deben considerar manuales técnicos, manuales de funcionamiento, manuales de usuario, manual de procedimientos específicos, etc.
- Especificar un cronograma de tiempo de puesta en marcha del proyecto, aquí se deberá considerar el tiempo que deberá tomar el proyecto en la fase de pruebas, se tomará en cuenta las sugerencias de los usuarios en cuanto a diseño de interfaz, colores, etc.
- El personal evaluador deberá emitir un informe donde verificará la garantía de las actividades de mayor prioridad del proyecto, con la finalidad de que el proyecto pase a la siguiente fase de pruebas, en el informe se deberá hacer constar diferentes parámetros de evaluación para garantizar la factibilidad del producto.

4.5 PRUEBAS

La finalidad de las pruebas es poder verificar el correcto funcionamiento de las tareas por las cuales se diseñó la aplicación y también para validar las expectativas del usuario, en las pruebas se deben ejecutar distintas tareas tomando en cuenta la funcionalidad del proyecto, los componentes de la plataforma seleccionada y los resultados obtenidos, estas pruebas son realizadas por el personal encargado de supervisar los resultados obtenidos, la ejecución de las pruebas debe cumplir con los alcances y limitaciones del proyecto planteado.

Este paso previo sirve para comprobar la validez del proyecto, la ejecución completa de las pruebas de los componentes de la plataforma de business intelligence ya que se necesita que el datamart contenga información y que los reportes de visualización estén funcionando correctamente, para esto es necesario cumplir los siguientes requerimientos:

- Verificación del datamart
- Pruebas de ETL
- Procedimientos ETL probados
- Diseño del cubo aprobado
- Pruebas de reportes
- Pruebas de funcionamiento

Para las pruebas del procedimiento del ETL de las fuentes de datos y carga del datamart usando herramientas acordadas para este proceso, se obtendrá como resultado un perfilamiento específico de los requerimientos solicitados. Para verificar el correcto funcionamiento de los reportes también es necesario realizar pruebas que cumplan con las expectativas de los usuarios, deberán estar basadas en los requerimientos del usuario.

4.6 RESULTADOS DE LAS PRUEBAS OBTENIDAS

Luego de haber analizado y aplicado el proceso de pruebas se obtienen los respectivos resultados, es muy importante que el sistema o prototipo cumpla con los requerimientos que exigen al hacer las pruebas, con estos resultados el encargado de realizar pruebas y mantenimiento del proyecto realiza un informe de posibles fallas o incumplimiento de parámetros, es para esto que se emite un informe final con varias recomendaciones y correcciones importantes con el objetivo de que cumplan con los objetivos planteados al inicio del proyecto, de esta manera se comprueba si hubo una perfecta coordinación entre el equipo de desarrollo y el equipo que levantó los requerimientos del proyecto, aportando así a la evaluación de pruebas objetividad y seriedad en este proceso, para concluir si el prototipo o proyecto cumple o no con las expectativas planteadas al inicio.

Dependiendo del tipo de evaluación que haga el encargado de monitorear el sistema deberá de emitir su informe con el porcentaje respectivo de cumplimiento de los parámetros propuestos anteriormente en las pruebas como resultado deberá emitir un informe con porcentajes de evaluación y recomendaciones o correcciones a ejecutar por el personal encargado de desarrollo.

4.7 MANUAL DE USUARIO

Ver Anexo A

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- A lo largo del desarrollo del presente trabajo de grado se ha podido observar que, la implementación de datamart con herramientas de software libre es muy confiable y seguro, obteniendo estabilidad al momento de integrar los datos, logrando buenos resultados.
- La creación de datamarts es un paso previo a la implementación de datawarehouse, conforme la organización o empresa va creciendo los datamarts van aumentando, permitiendo toma de decisiones gerenciales de acuerdo al área de negocio o departamento que necesita mayor énfasis en ser administrado.
- La implementación de business intelligence en una organización, pequeña o mediana hace que la información permanezca ordenada y controlada, registrando las transacciones diarias que realiza la empresa hasta convertirse en información histórica valiosa para ser analizada.
- Para tener una buena implementación de inteligencia de negocios es muy importante conocer bien los requerimientos académicos del área de negocio y hacer un análisis completo de la información que va a ser analizada.
- La implementación de datawarehouse en grandes volúmenes de información resulta un proceso largo pero esencial cuando se necesita hacer análisis de información, ya que permite agrupar y clasificar en grupos significativos la información más importante.
- Los beneficios que se obtienen al realizar la comparación de tres herramientas de software libre, resulta muy provechoso a la hora de implementar este tipo de proyectos, con la demostración de la herramienta seleccionada, el correcto funcionamiento, la fácil utilización y depuración de la misma.

- De los resultados obtenidos se destaca la valiosa experiencia obtenida en el área de Inteligencia de Negocios en el transcurso del desarrollo del presente prototipo.
- Se investigó las herramientas de software libre Birt, Attivio y Pentaho lo que permitió determinar a lo largo del desarrollo del presente trabajo de grado que la Suite de Pentaho, tiene muchas más funcionalidades y es una herramienta confiable para la elaboración de datamarts.
- La suite de Pentaho permitió la elaboración de datamart con datos reales con el objeto de demostrar la el tema planteado en el presente trabajo de grado.
- La herramienta Pentaho Reporting permitió la elaboración de reportes luego de haber realizado el cubo de información.

5.2 RECOMENDACIONES

- Para una exitosa implementación y para futuras adiciones es necesario trabajar en equipo tanto las áreas de negocio involucradas como el departamento de tecnología de Información.
- Es muy importante tomar en cuenta la arquitectura y los componentes de software a utilizar, su compatibilidad y unificación con otras plataformas porque pueden existir futuras adiciones o migración a una nueva plataforma.
- Al momento de haber terminado la implementación se recomienda dar seguimiento a la información que se carga en el datamart con la finalidad de mantener información actualizada y confiable que sea de beneficio para los usuarios que está encargados en la toma de decisiones.
- Las asignaciones de roles para las distintas áreas del negocio deben ser definidas por el personal encargado de monitorear la información.
- Es recomendable utilizar herramientas de software libre en las pequeñas y medianas empresas ya que permiten reducir costos, ahorrar recursos económicos, la suite Pentaho es una solución que nos permite la gestión y análisis de información que puede ser utilizada por cualquier organización que desee implementar una solución de Inteligencia de Negocios.
- La correcta identificación de los requerimientos y necesidades se lo debe hacer directamente con los usuarios y gerentes de las organizaciones a implementar Inteligencia de Negocios con el objetivo de que la información sea segura y confiable para evitar futuros problemas en el desarrollo del sistema.
- Se recomienda a los docentes incentivar el uso de herramientas de software libre que faciliten el desarrollo de proyectos de software.
- Se recomienda realizar estudios comparativos con herramientas de software libre que permitan la elaboración de minería de datos.

GLOSARIO DE TÉRMINOS

- **Actuate Corporation.-** Empresa con sede en San Mateo, California, es conocida por su creación del código abierto Eclipse BIRT puesto en marcha por la Fundación Eclipse en 2004.
- **Add-On.-** Conocidos como extensiones, plugins, snap-ins, etc., son programas que sólo funcionan anexados a otro y que sirven para incrementar o complementar sus funcionalidades.
- **AIE.-** Active Intelligence Engine.
- **Alfresco.-** Es un sistema de administración de contenidos de código fuente libre, desarrollado en Java, basado en estándares abiertos y de escala empresarial para sistemas operativos tipo Windows, Unix Solaris y algunas versiones de Linux.
- **Apache Hadoop.-** Apache Hadoop es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre.
- **Back-end.-** Trabaja conjuntamente con el front-end que es el responsable de recolectar los datos de entrada del usuario, que pueden ser de muchas y variadas formas, y los transforma ajustándolos a las especificaciones que demanda el back-end para poder procesarlos, devolviendo generalmente una respuesta que el front-end recibe y expone al usuario de una forma entendible para este. La conexión del front-end y el back-end es un tipo de interfaz.
- **Bill Inmon.-** Reconocido como el padre del almacén de datos fue el primero en ofrecer clases en el almacenamiento de datos. Bill Inmon creó la definición aceptada de lo que es un almacén de datos: un tema orientado, no volátil, variante colección integrada, el tiempo de los datos en apoyo de las decisiones de la gerencia.
- **Business intelligence.-** Inteligencia de negocios.
- **Cidescol.-** es un corredor asesor de seguros con una reconocida trayectoria de seriedad y servicio trabajando con las principales y más sólida compañía.

- **CMI.-** Cuadros de Mando Integrales.
- **Commit.-** Sentencia en SQL que finaliza una transacción de base de datos dentro de un sistema gestor de base de datos relacional (RDBMS) y pone visibles todos los cambios a otros usuarios.
- **Clksoft Cía. Ltda.-** Es una empresa enfocada en implementar soluciones de business intelligence y visualización de Información para mejorar el proceso de toma de decisiones en sus clientes.
- **Cloud computing.-** La computación en la nube son servidores desde Internet encargados de atender las peticiones en cualquier momento. Se puede tener acceso a su información o servicio, mediante una conexión a internet desde cualquier dispositivo móvil o fijo ubicado en cualquier lugar.
- **Clustering:** Los algoritmos de clustering son usualmente empleados para mejorar el rendimiento y/o la disponibilidad de la información en una minería de datos.
- **CRM.-** Customer Relationship Management que significa: Gestión de Relaciones con Clientes.
- **Dashboards:** Cuadros de mandos de control.
- **Dreamweaver.-** Es una aplicación en programa de estudio que está destinada a la construcción, diseño y edición de sitios, vídeos y aplicaciones Web basados en estándares.
- **Drilling.-** Significa navegación entre arquitecturas que puede ser horizontal o vertical.
- **DSS.-** Sistemas de Soporte a la Decisión.
- **Eclipse Foundation.-** La Fundación Eclipse lidera el desarrollo de Eclipse, el IDE y plataforma de código abierto para el desarrollo de aplicaciones en Java.
- **EIS.-** Sistemas de Información Ejecutiva.

- **ERP.-** Enterprise Resource Planning, sistemas informáticos destinados a la administración de recursos en una organización.
- **ETL.-** Extracción, transformación y carga.
- **FICA.-** Facultad de Ingeniería en Ciencias Aplicadas.
- **Fybeca.-** Fybeca es una empresa orgullosamente ecuatoriana, pionera en el sector de comercio al detalle farmacéutico.
- **IBM.-** International Business Machines Corp. es una empresa multinacional estadounidense de tecnología y consultoría, fabrica y comercializa hardware y software para computadoras, ofrece servicios de infraestructura, alojamiento de Internet y consultoría en una amplia gama de áreas relacionadas con la informática, desde computadoras centrales hasta nanotecnología.
- **Innovent Solutions.-** Es una empresa de consultoría, dedicada a la mejora de las decisiones a través de la gestión eficaz de la información de decisiones.
- **JasperReports.-** Es una librería de creación de informes que tiene la habilidad de entregar contenido enriquecido al monitor, a la impresora o a ficheros.
- **Javascript.-** Es un lenguaje de programación interpretado, dialecto del estándar ECMAScript. Se define como orientado a objetos, basado en prototipos, imperativo, débilmente tipado y dinámico.
- **JDBC.-** Es una API que permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java, independientemente del sistema operativo donde se ejecute o de la base de datos a la cual se accede, utilizando el dialecto SQL del modelo de base de datos que se utilice.
- **JfreeReport.-** Es una biblioteca de informe libre de Java.
- **JNDI.-** Es una Interfaz de Programación de Aplicaciones (API) de Java para servicios de directorio. Permite a los clientes descubrir y buscar objetos y datos a través de un nombre.

- **JSP.**- Es una tecnología que ayuda a los desarrolladores de software a crear páginas web dinámicas basadas en HTML, XML, entre otros tipos de documentos.
- **MDX:** Es el lenguaje con el que trabaja el servidor OLAP Mondrian.
- **OpenBravo.**- Es una aplicación de gestión empresarial del tipo ERP destinada a empresas de pequeño y mediano tamaño.
- **OpenL.**- Es un marco innovador para el desarrollo de las diferentes configuraciones de idioma, es una metodología y herramientas de apoyo que permiten a los programadores profesionales para crear sus propias versiones de los lenguajes.
- **Open Source.**- Código abierto es la expresión con la que se conoce al software distribuido y desarrollado libremente.
- **Plugin.**- Es una aplicación que se relaciona con otra para aportarle una función nueva y generalmente muy específica.
- **QuikView.**- Capacita a los usuarios de negocio para que tomen decisiones de forma innovadora.
- **Queries.**- Es una consulta o una búsqueda de datos almacenados en una base de datos.
- **Ralph Kimball.**- Él es ampliamente considerado como uno de los arquitectos originales de almacenamiento de datos y es conocido por convicciones a largo plazo que los almacenes de datos deben estar diseñados para ser comprensible y rápido. Su metodología, también conocido como el modelado tridimensional, se ha convertido en el estándar de facto en el área de apoyo a la decisión.
- **Renderizado.**- Se refiere al proceso de generar una imagen o vídeo mediante el cálculo de iluminación GI partiendo de un modelo en 3D.

- **Rollback.**- Operación que devuelve a la base de datos a algún estado previo. Los Rollbacks son importantes para la integridad de la base de datos, la base de datos puede ser restaurada a una copia limpia incluso después de que se han realizado operaciones erróneas.
- **RUP.**- Es un proceso de desarrollo de software desarrollado por la empresa Rational Software, actualmente propiedad de IBM.
- **Servlets.**- Es una clase en el lenguaje de programación Java, utilizada para ampliar las capacidades de un servidor.
- **SSOO.**- Sistema Operativo.
- **Stakeholders:** Son las personas interesadas o involucradas en un proyecto determinado, y que necesitan generar una solución.
- **Sumarización.**- Muestra los datos de una manera más resumida, permitiendo, precisamente, calcular valores agregados, que no son los datos directos registrados, sino datos derivados de ellos.
- **Talend.**- Es un código abierto que proporciona el proveedor de software de integración, gestión de datos, integración de aplicaciones empresariales y datos de gran software y servicios.
- **Tomcat.**- funciona como un contenedor de servlets desarrollado bajo el proyecto Jakarta en la Apache Software Foundation.
- **Top-down.**- Formula un resumen del sistema, sin especificar detalles, es una estrategia de procesamiento de información.
- **Url.**- Está formado por una secuencia de caracteres, de acuerdo a un formato modélico y estándar, que designa recursos en una red, como Internet.
- **Weka:** Es una plataforma de software para aprendizaje automático y minería de datos escrito en Java y desarrollado en la Universidad de Waikato, Weka es un software libre distribuido bajo licencia GNU-GPL.

- **War.**- Es un tipo de archivo que permite empaquetar una aplicación web para que sea desplegada en un contenedor web JEE, o en un servidor de aplicaciones JEE.
- **Wizard:** Asistente de ayuda.
- **Workflow:** O flujo de trabajo es el estudio de los aspectos operacionales de una actividad de trabajo: cómo se estructuran las tareas, cómo se realizan, cuál es su orden correlativo, cómo se sincronizan, cómo fluye la información que soporta las tareas y cómo se le hace seguimiento al cumplimiento de las tareas.

BIBLIOGRAFÍA

- Kimball Ralph. (2008). *The Data warehouse Life Cycle Toolkit*. Indiana: Wiley.
- Wrembel Robert. (2007). *Data warehouses and OLAP: concepts, architectures and solutions*. Estados Unidos: IRM Press.

LINCOGRAFÍA

- Antonio Rivas. (23 de Febrero de 2011). *Análisis del cuadrante de Gartner – Business Intelligence 2011. Parte 3*. Obtenido de <http://www.bi.dev42.es/2011/02/23/olap-molap-rolap/>.
- Attivio. (16 de Octubre de 2013). *Company Overview*. Obtenido de <http://www.crunchbase.com/organization/attivio>.
- Attivio. Inc., (20 de Junio de 2014). *Complete Unified Information Access for Business Innovation*. Obtenido de <http://www.attivio.com/how-we-do-it/the-platform/active-intelligence-engine>.
- Christopher Nuzum. (17 de Noviembre de 2010). *Attivio Search System Requirements*. Obtenido de <https://support.tractionsoftware.com/traction#/single&proj=AttivioSearch&rec=4&brief=n>.
- Deinsa. (25 de Abril de 2014). *Inteligencia de Negocios*. Obtenido de <http://www.idensa.com/>
- EOS. (25 de Febrero de 2014). *Business Intelligence QlickView*. Obtenido de <http://www.eopensolutions.com/productos-y-servicios/business-intelligence-qlikview-ecuador>
- fabFORCE.net. (25 de Enero de 2014). *DBDesigner4*. Obtenido de <http://www.fabforce.net/dbdesigner4/>
- Hide Julian. (23 de Agosto de 2014). *Mondrian Documentation*. Obtenido de http://mondrian.pentaho.com/documentation/installation_es.php

- Honorable Consejo Universitario. (12 de Noviembre de 2005). *REGLAMENTO GENERAL - UTN*. Obtenido de <http://www.utn.edu.ec/web/portal/images/doc-utn/reglamento-general-utn.pdf>
- Kimball y Wrembel. (2 de Abril de 2011). *Los datos de transacciones crecen*. Obtenido de <http://www.informatica.com/es/vision/harnessing-big-data/oltp-and-analytics/>.
- Kimball, R. (2008). *The Data warehouse Life Cycle Toolkit*. Indiana: Wiley.
- Mariana Isabel Rojas. (20 de Junio de 2009). *Data warehouse*. Obtenido de <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonoAdsDiseno.pdf>.
- Marysol Tamayo, Javier Moreno. (20 de Junio de 2010). *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP*. Obtenido de http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci_arttext.
- Meteorite Consulting. (16 de Enero de 2014). *Saiku Enterprise Analytics*. Obtenido de <http://www.meteorite.bi/>
- Microsoft. (02 de Enero de 2013). *Procesamiento y modos de almacenamiento de particiones*. Obtenido de <http://msdn.microsoft.com/es-es/library/ms174915.aspx>.
- Roberto Espinosa. (19 abril de 2010). *Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional*. Obtenido de <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimencional/>.
- Sinnexus. (20 de Noviembre de 2012). *Datawarehouse*. Obtenido de http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
- Sinnexus. (21 de Enero de 2014). *¿Qué es Business Intelligence?* Obtenido de http://www.sinnexus.com/business_intelligence/

- Thomas. (03 de Julio de 2014). *Reporting Tales*. Obtenido de <https://www.on-reporting.com/what-is-pentaho-reporting/>
- Wikipedia. (26 de 04 de 2013). *HOLAP*. Obtenido de <http://es.wikipedia.org/wiki/HOLAP>
- Wikipedia. (17 de 10 de 2013). *OLTP*. Obtenido de <http://es.wikipedia.org/wiki/OLTP>
- Wikipedia. (05 de Noviembre de 2013). *ROLAP*. Obtenido de <http://es.wikipedia.org/wiki/ROLAP>
- Wikipedia. (20 de Marzo de 2014). *Almacén de Datos*. Obtenido de http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos
- Wikipedia. (13 de Noviembre de 2014). *Attivio*. Obtenido de <http://en.wikipedia.org/wiki/Attivio>
- Wikipedia. (26 de Marzo de 2014). *Business Intelligence and Reporting Tools*. Obtenido de http://es.wikipedia.org/wiki/Business_Intelligence_and_Reporting_Tools
- Wikipedia. (15 de Noviembre de 2014). *Datamining*. Obtenido de http://en.wikipedia.org/wiki/Data_mining
- Wikipedia. (17 de Junio de 2014). *Granularidad*. Obtenido de <http://es.wikipedia.org/wiki/Granularidad>
- Wikipedia. (21 de 06 de 2014). *MOLAP*. Obtenido de <http://es.wikipedia.org/wiki/MOLAP>
- Wikipedia. (05 de 06 de 2014). *OLAP*. Obtenido de <http://es.wikipedia.org/wiki/OLAP>
- Wikipedia. (16 de Julio de 2014). *Pentaho*. Obtenido de <http://es.wikipedia.org/wiki/Pentaho>

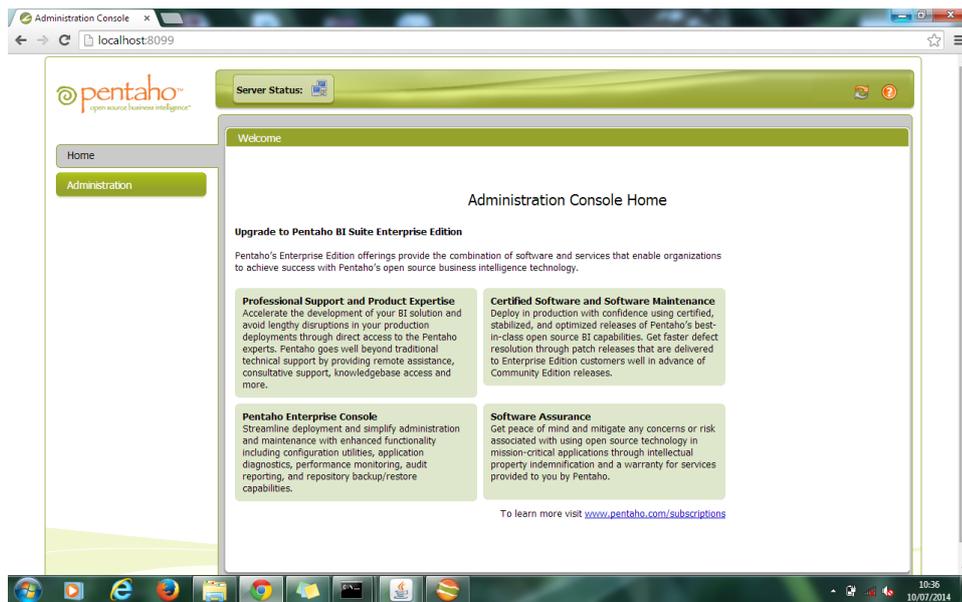
- Wikipedia. (18 de Junio de 2014). *Sistema de planificación de recursos empresariales*. Obtenido de http://es.wikipedia.org/wiki/Sistema_de_planificaci%C3%B3n_de_recursos_empresariales.
- Wood Sherman. (27 de Abril de 2014). *Documentación Mondrian*. Obtenido de <http://mondrian.pentaho.com/documentation/workbench.php>

ANEXOS

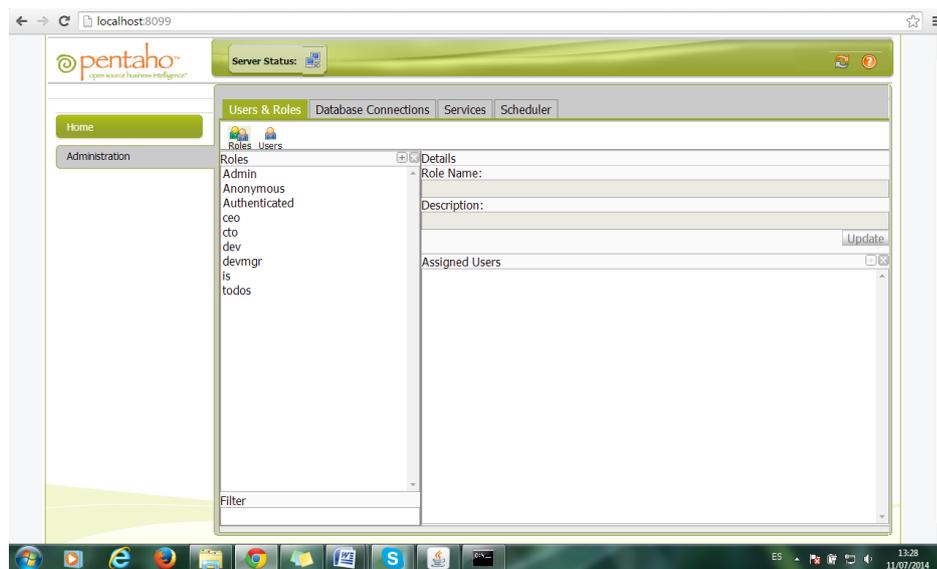
ANEXO A

MANUAL DE USUARIO PARA LA CREACIÓN DE USUARIOS Y ASIGNACIÓN DE ROLES Y PERMISOS EN EL SERVIDOR MONDRIAN

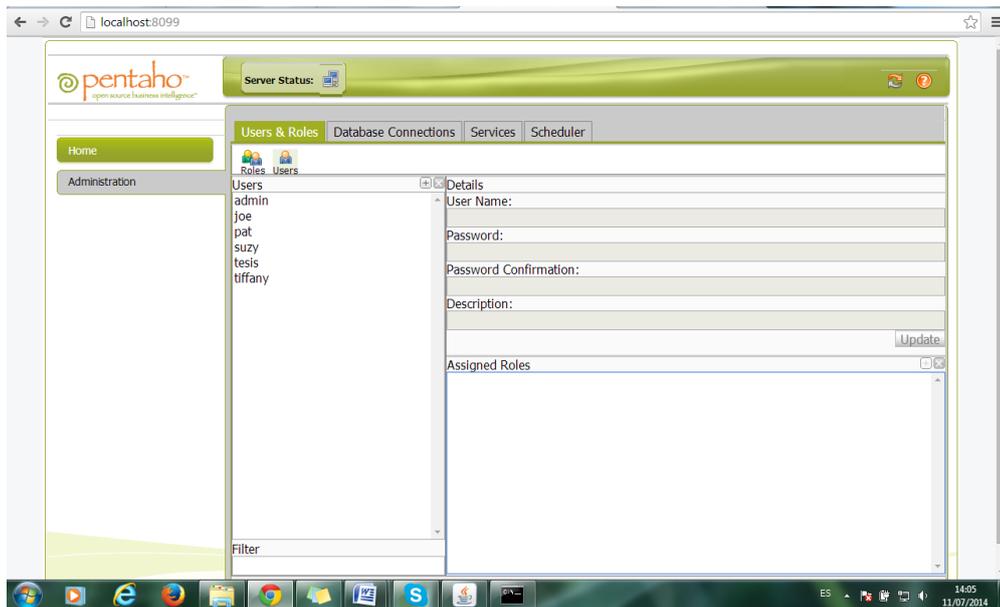
1. Deberá estar ubicado en la pantalla de administración de la consola del servidor Mondrian.



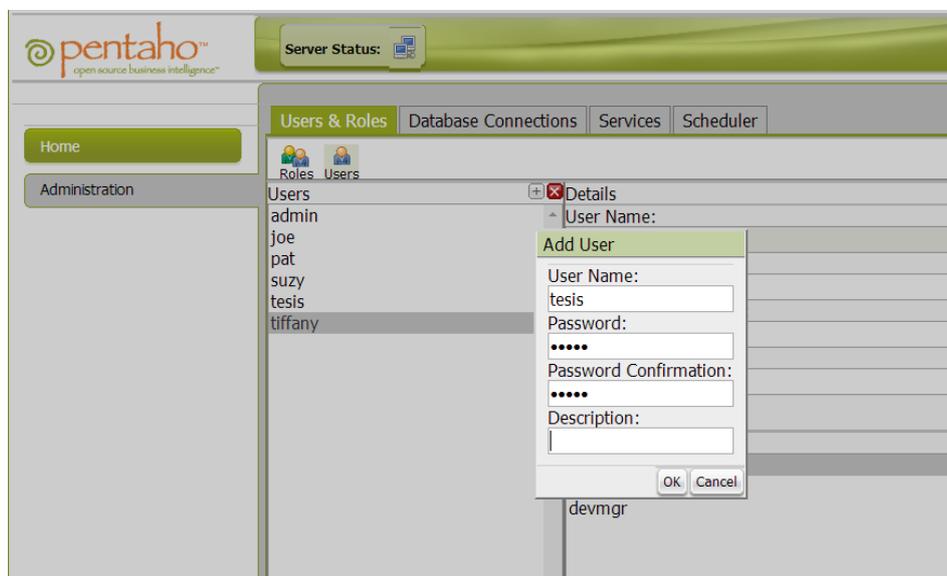
2. Dar clic en la pestaña **Administración**, se desplegará una pantalla con varias pestañas, deberá seleccionar la pestaña **Users & Roles**.



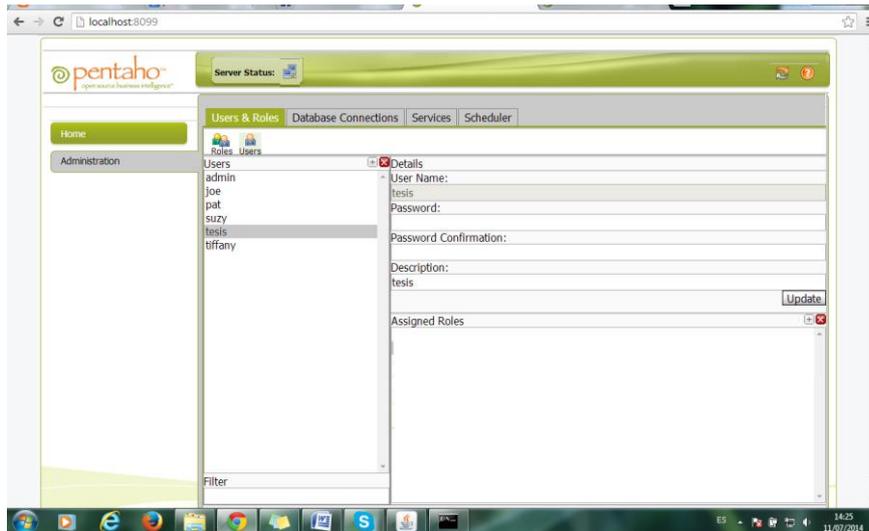
3. A continuación deberá dar clic en **Users**, se despliega una lista de usuarios que la herramienta tiene creado por defecto.



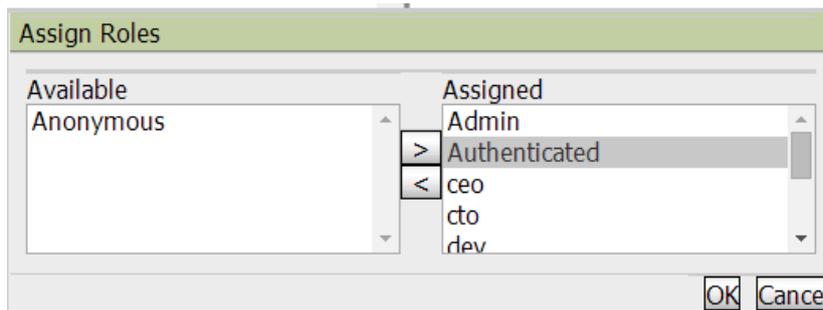
4. Para agregar un nuevo usuario deberá dar clic en el signo **MAS**, se desplegará una nueva ventana donde deberá ingresar el nombre del usuario a crear y la contraseña:



5. Para actualizar el nombre de usuario o contraseña del usuario creado solo deberá seleccionar el usuario a modificar y se desplegará una ventana una vez realizados los cambios deberá dar clic en **Update**.



- Para asignar roles al usuario creado deberá seleccionar el usuario que creó y en la parte derecha inferior deberá dar clic en el signo **MAS**, se desplegará una nueva ventana donde verá seleccionar el tipo de rol que deberá tener el usuario creado.



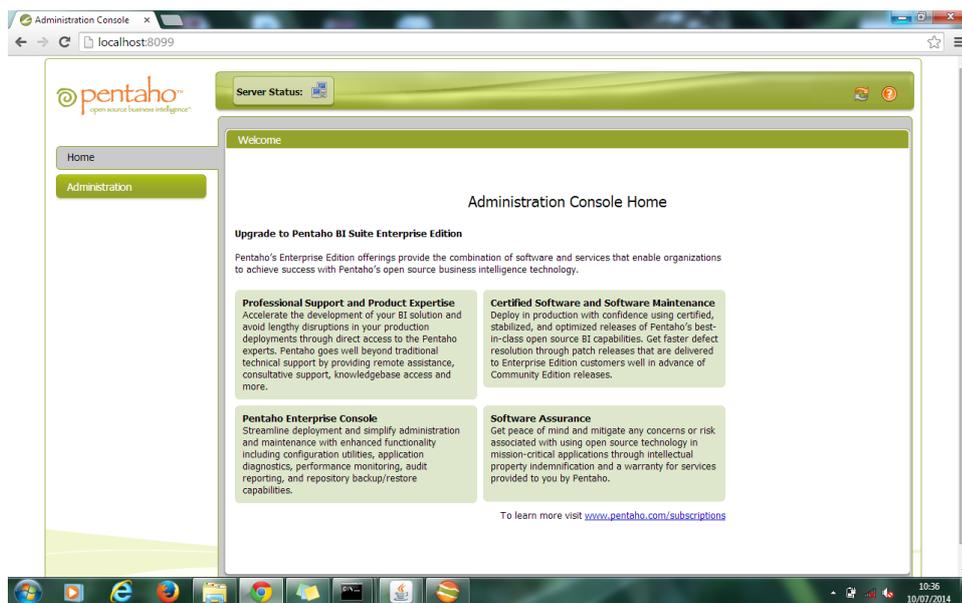
- Este usuario es el que le permitirá loguearse e ingresar a la página principal del servidor Mondrian para realizar los análisis en el cubo.



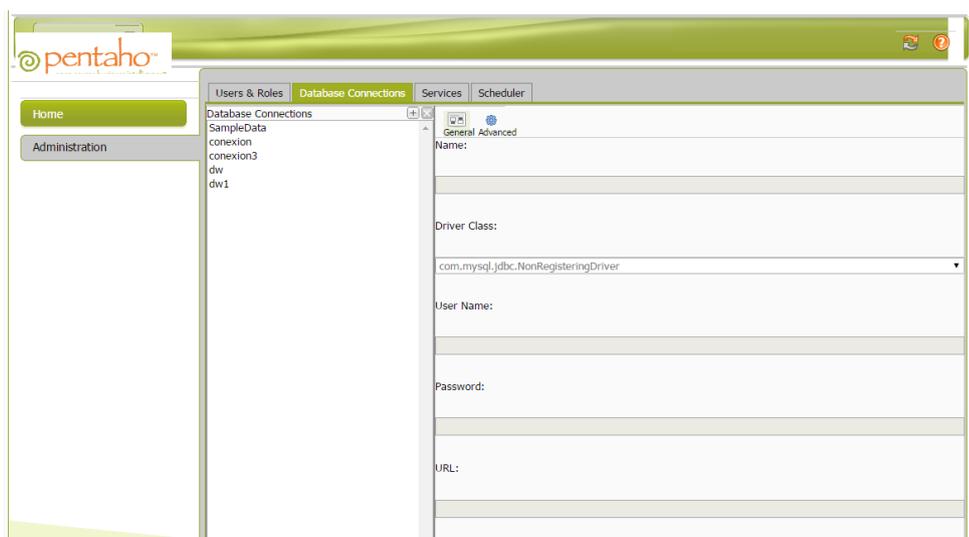
ANEXO B

MANUAL DE USUARIO PARA LA CREACIÓN DE UNA CONEXIÓN CON LA BASE DE DATOS EN EL SERVIDOR MONDRIAN PARA LA PUBLICACIÓN DE CUBOS

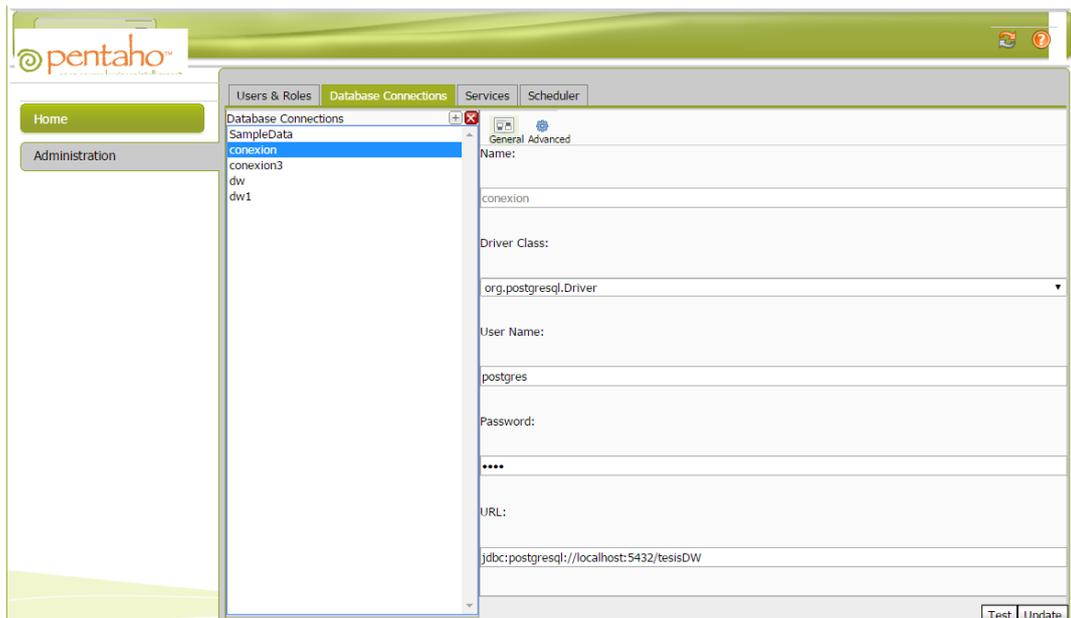
1. Deberá estar ubicado en la pantalla de administración de la consola del servidor Mondrian.



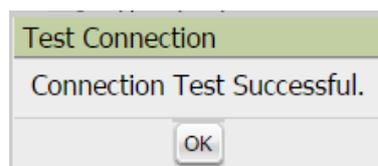
2. Dar clic en la pestaña **Database Connections**, se desplegará una pantalla con varias opciones a llenar, deberá completar todos los campos.



3. Deberá escoger un nombre para su conexión, en **Driver Class:** deberá escoger la librería correspondiente con la base de datos que está trabajando, en **User Name:** deberá poner el nombre de usuario de la base de datos, en **Password:** deberá poner la contraseña del usuario de la base de datos, y en **URL:** deberá poner la conexión JDBC de la base de datos.



4. Una vez completados los campos obligatorios, deberá dar clic en el botón **Test**, se desplegará una ventana comprobando la autenticidad de los datos.



5. Para actualizar algún dato en la conexión creada deberá seleccionar el nombre de la conexión a modificar y se desplegará una ventana con los datos llenos, una vez realizados los cambios deberá dar clic en **Update** y se desplegará una ventana de aceptación.

