



UNIVERSIDAD TECNOLÓGICA EQUINOCCIAL

**FACULTAD DE CIENCIAS DE LA INGENIERÍA E
INDUSTRIAS**

**CARRERA DE INGENIERÍA INFORMÁTICA Y
CIENCIAS DE LA COMPUTACIÓN**

**CREACIÓN DE UN DATA WAREHOUSE, DEFINIR EL
PROCESO DE INTEGRACIÓN CARGA DE DATOS Y
GENERAR REPORTES PARA TOMA DE DECISIONES POR
MEDIO DEL USO DE HERRAMIENTAS DE INTELIGENCIA DEL
NEGOCIO Y MINERÍA DE DATOS EDUCACIONALES PARA LA
GESTIÓN DEL CONOCIMIENTO DEL ÁREA DE
PLANIFICACIÓN DE LA UTE.**

**TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO
DE INGENIERO INFORMÁTICO Y CIENCIAS DE LA COMPUTACIÓN**

LUIS ANDRÉS SAMPEDRO GERMAN

DIRECTOR: ING. OSWALDO MOSCOSO, MSC

Quito, Mayo 2016

© Universidad Tecnológica Equinoccial. 2016
Reservados todos los derechos de reproducción

DECLARACIÓN

Yo **LUIS ANDRÉS SAMPEDRO GERMAN**, declaro 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.

La Universidad Tecnológica Equinoccial puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normativa institucional vigente.

A rectangular image showing a handwritten signature in blue ink on a light-colored background. The signature is stylized and appears to read 'Luis Andrés Sampedro German'.

LUIS ANDRÉS SAMPEDRO GERMAN

C.I.1721624771

CERTIFICACIÓN

Certifico que el presente trabajo que lleva por título **“Creación de un data warehouse, definir el proceso de integración carga de datos y generar reportes para toma de decisiones por medio del uso de herramientas de inteligencia del negocio y minería de datos educacionales para la gestión del conocimiento del área de planificación de la UTE”**, que, para aspirar al título de **Ingeniero Informático y Ciencias de la Computación** fue desarrollado por **Andrés Sampedro**, bajo mi dirección y supervisión, en la Facultad de Ciencias de la Ingeniería; y cumple con las condiciones requeridas por el reglamento de Trabajos de Titulación artículos 18 y 25.



Msc. Oswaldo Moscoso

DIRECTOR DEL TRABAJO

C.I.1709629651

DEDICATORIA

El proyecto de titulación está dedicado especialmente a mi familia, mis padres Luis y Blanca, mis hermanos Diego y Stefany que han sido un pilar fundamental en mi vida y a pesar de todas las adversidades y problemas que se suscitaron en el camino siempre han confiado en mí y me han apoyado y motivado en cada momento, dándome mucho cariño y fuerzas para seguir adelante en la consecución de este proyecto. Gracias a mi familia por su ardua labor de inculcarme los valores necesarios que sirvieron como fuente de inspiración día a día para formarme como persona, y de aquí en adelante en mi vida profesional, como a mis abuelitos Hermógenes, Rosita y Luz María que con su preocupación, ejemplo y cariño siempre me daban ánimos en toda mi vida universitaria, a mis queridos primos Alex, Paola, Santiago, Luis, Álvaro, que han sido un gran apoyo desde que comencé con mi carrera, como olvidar a mis queridos amigos que estuvieron en este camino que no fue fácil, los quiero y aprecio mucho Santiago, Bryan, Víctor, Dennis, Jonathan, Fernando, Jorge, Carlos, Ángel, Katty, Vivi, Cristina, Verónica, Thalía, Belén, Naty, gracias por brindarme su estímulo incondicional en cada momento como a un hermano y hacerme ver que nunca estuve solo. Todos y cada uno contribuyó desde el comienzo de mi carrera hasta la consecución del proyecto y les estaré agradecido por siempre. Gracias por ayudarme a alcanzar mi meta.

“Si tú no trabajas por tus sueños, alguien te contratará para que trabajes por los suyos”

(Steve Jobs, Cofundador de Apple)

AGRADECIMIENTO

Agradezco primeramente a Dios por iluminarme y guiarme en cada paso que he dado en mi vida y darme una maravillosa familia.

A todos los mencionados anteriormente en mi dedicatoria, A mis profesores que conforman la carrera de Ingeniería Informática de la Universidad Tecnológica Equinoccial que a lo largo de mi carrera universitaria me han guiado y me brindados sus conocimientos.

Bueno todos y cada uno de ustedes me han acompañado y han plasmado en mí una parte de su ser, de verdad gracias, de ustedes he aprendido mucho, me han marcado porque me han enseñado cosas que los libros no enseñan, porque gracias a ustedes puedo decir que voy en camino para ser un gran hombre, que siempre busque el bien para los demás.

FORMULARIO DE REGISTRO BIBLIOGRÁFICO

PROYECTO DE TITULACIÓN

DATOS DE CONTACTO	
CÉDULA DE IDENTIDAD:	1721624771
APELLIDO Y NOMBRES:	SAMPEDRO GERMAN LUIS ANDRÉS
DIRECCIÓN:	SABANILLA OE4-672 Y AV. LA PRENSA
EMAIL:	landres_7@hotmail.com
TELÉFONO FIJO:	2533485
TELÉFONO MOVIL:	0999879627

DATOS DE LA OBRA	
TÍTULO:	CREACIÓN DE UN DATA WAREHOUSE, DEFINIR EL PROCESO DE INTEGRACIÓN CARGA DE DATOS Y GENERAR REPORTES PARA TOMA DE DECISIONES POR MEDIO DEL USO DE HERRAMIENTAS DE INTELIGENCIA DEL NEGOCIO Y MINERÍA DE DATOS EDUCACIONALES PARA LA GESTIÓN DEL CONOCIMIENTO DEL ÁREA DE PLANIFICACIÓN DE LA UTE
AUTOR O AUTORES:	SAMPEDRO GERMAN LUIS ANDRÉS
FECHA DE ENTREGA DEL PROYECTO DE TITULACIÓN:	2016-04-16
DIRECTOR DEL PROYECTO DE TITULACIÓN:	OSWALDO MOSCOSO
PROGRAMA	PREGRADO <input checked="" type="checkbox"/> POSGRADO <input type="checkbox"/>

TÍTULO POR EL QUE OPTA:	INGENIERO INFORMÁTICO Y CIENCIAS DE LA COMPUTACIÓN
RESUMEN: Mínimo 250 palabras	<p>La información se ha convertido en un elemento de gran importancia para el desarrollo de las instituciones por lo cual es importante conocer el ambiente en el que se desenvuelven para tomar acertadas decisiones. Por esta razón los datos históricos y las bases de datos operacionales depuradas y unificadas son fuente de información valiosa para cada institución.</p> <p>El proyecto de titulación está orientado a proporcionar una herramienta informática que ayude con la toma de decisiones en el departamento de planificación. Este departamento es encargado de establecer los diferentes planes para el proceso de acreditación institucional, proceso que es realizado cada cinco años por el CEAACES y se basa en la eficiencia académica, investigación, organización e infraestructura.</p> <p>Tomando en cuenta la problemática mencionada anteriormente, para el presente proyecto se desarrolla un <i>data warehouse</i> con herramientas de inteligencia de negocios y minería de datos para facilitar la toma de decisiones oportunas ayudando a identificar debilidades en los procesos de la institución y rectificarlos de una manera eficiente, convirtiéndose en fortalezas para la universidad. Para el proceso de desarrollo del <i>data warehouse</i> se estableció como metodología la de <i>Ralph Kimball</i> permitiendo una planificación desde el inicio hasta el final de proyecto. Esta propuesta incluye un</p>



	<p>modelo de gestión del conocimiento y los diferentes requisitos que necesita el departamento para que el <i>data warehouse</i> pueda tener un mantenimiento y un crecimiento posterior.</p> <p>Para el aprovechamiento de esta herramienta se utilizó <i>Microsoft Excel</i> con complementos de minería de datos que facilitan la creación de reportes dinámicos y el análisis de modelos predictivos en busca de tendencias desconocidas que apoyen al proceso académico e institucional.</p>
PALABRAS CLAVES:	Data warehouse, Minería de datos, Reportes, Dashboard, CEAACES, Kimball.
ABSTRACT:	<p>The use of information in time has become a relevant element in the development of enterprises and institutions. Information helps people to recognize the internal and external advantages and disadvantages and to understand the environment in which they operate. The main objective is to visualize mistakes, evaluate and make good decisions that will improve the performance of organizations. For this reason, the use of historical facts and operational databases are valuable sources of knowledge.</p> <p>The objective of this academic project is to implement an informatic tool that supports the decision making process in the planning department of the university. This department is in charge of developing strategies and plans that contribute to the</p>



	<p>institutional process of accreditation by CEAACES. This process is done every five years, and is based on academic efficiency, research development, and the correct organization and use of infrastructure.</p> <p>Taking into account this problem, in this project a data warehouse with business intelligence and data mining tools was developed to facilitate proper decision making, to identify weaknesses in institutional processes and to propose solutions that are viable and appropriate to strengthen the university. For the development process of the data warehouse a knowledge management model was used which takes into account the requirements established in Ralph Kimball methodology. This methodology supports the planning process from the beginning to the end of the project, which is reflected in a proposal that takes into account the requirements of the planning department for the proper operation of the tool and its maintenance over time.</p> <p>Data mining complements of Microsoft Excel are used for the efficient job and exploitation of the data warehouse. This facilitate the creation of dynamic reports and analysis of predictive models and the discovery of unknown trends that contributes to the academic and the institutional process.</p>
KEYWORDS	Data warehouse, Data Mining, Reports, dashboard, Kimball, CEAACES.

Se autoriza la publicación de este Proyecto de Titulación en el Repositorio Digital de la Institución.



SAMPEDRO GERMAN LUIS ANDRÉS

1721624771

DECLARACIÓN Y AUTORIZACIÓN

Yo, **SAMPEDRO GERMAN LUIS ANDRÉS**, CI 1721624771 autor/a del proyecto titulado: **Creación de un Data Warehouse, definir el proceso de integración carga de datos y generar reportes para toma de decisiones por medio del uso de herramientas de inteligencia del negocio y minería de datos educacionales para la gestión del conocimiento del área de Planificación de la UTE**. Previo a la obtención del título de **INGENIERO EN INFORMÁTICA Y CIENCIAS DE LA COMPUTACIÓN** en la Universidad Tecnológica Equinoccial.

1. Declaro tener pleno conocimiento de la obligación que tienen las Instituciones de Educación Superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de graduación para que sea integrado al Sistema Nacional de información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.
2. Autorizo a la BIBLIOTECA de la Universidad Tecnológica Equinoccial a tener una copia del referido trabajo de graduación con el propósito de generar un Repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Quito, 10 de Mayo del 2016



SAMPEDRO GERMAN LUIS ANDRÉS

1721624771

TABLA DE CONTENIDOS

	PÁGINA
RESUMEN	xi
ABSTRACT	xii
1. INTRODUCCIÓN.....	1
2. MARCO TEÓRICO	5
2.1. GESTIÓN DEL CONOCIMIENTO	5
2.1.1. Dato.....	5
2.1.2. Información	5
2.1.3. Conocimiento	5
2.1.4. Conocimiento explícito	6
2.1.5. Conocimiento tácito.....	6
2.1.6. Gestión del conocimiento	6
2.1.7. Necesidades de Información	7
2.1.8. Fuentes de generación de conocimiento.....	7
2.1.8.1. Descubrimiento del conocimiento en bases de datos	9
2.1.9. Modelos para la gestión del conocimiento	10
2.1.9.1. Modelo de creación de conocimiento organizacional Nonaka y Takeuchi	10
2.1.9.2. Modelo de creación de conocimiento Technology Broker.....	12
2.2. INTELIGENCIA DE NEGOCIOS.....	13
2.2.1. Definición de inteligencia de negocios	13
2.2.2. Arquitectura de inteligencia de negocios.....	13
2.2.2.1. Fuentes de datos	14
2.2.2.2. Integración de datos	14
2.2.2.3. Análisis de datos.....	14
2.2.2.4. Reportes	15
2.2.2.5. Usuario.....	15
2.2.3. Componentes.....	15
2.2.3.1. Diseño conceptual de los sistemas	15
2.2.3.2. Construcción y alimentación del data warehouse	15
2.2.3.3. Herramientas de explotación del data warehouse	16

2.2.3.3.1.	Consultas y reportes.....	16
2.2.3.3.2.	Cuadro de mando analítico.....	16
2.2.3.3.3.	Cuadro de mando integral	16
2.3.	DATA WAREHOUSE.....	16
2.3.1.	Definición de data warehouse	16
2.3.2.	Características un data warehouse	17
2.3.2.1.	Orientado al tema	17
2.3.2.2.	Integración	17
2.3.2.3.	De tiempo variante.....	17
2.3.2.4.	No volátil	17
2.3.3.	Arquitectura de un data warehouse.....	18
2.3.3.1.	Componentes de un data warehouse	18
2.3.3.1.1.	Datos operacionales	18
2.3.3.1.2.	Datos históricos	18
2.3.3.1.3.	Gestor de carga.....	19
2.3.3.1.4.	Administrador del data warehouse	19
2.3.3.1.5.	Gestor de consultas.....	19
2.3.3.1.6.	Datos detallados.....	19
2.3.3.1.7.	Metadatos.....	19
2.3.3.1.8.	Herramientas de minería de datos.....	20
2.4.	MINERÍA DE DATOS.....	20
2.4.1.	Definición de minería de datos.....	20
2.4.2.	Técnicas de minería de datos	20
2.4.2.1.	Modelo predictivo o de aprendizaje supervisado	20
2.4.2.1.1.	Modelo de clasificación.....	21
2.4.2.1.1.1.	Arboles de decisión	21
2.4.2.1.1.2.	Redes neuronales	21
2.4.2.1.1.3.	Redes bayesianas	21
2.4.2.2.	Modelo descriptivo o de aprendizaje no supervisado	21
2.4.2.2.1.1.	Segmentación.....	22
2.4.2.2.1.2.	Detección de desviación.....	22
2.4.3.	Minería de datos educacionales.....	22

2.5.	RALPH KIMBALL.....	23
2.5.1.	Arquitectura.....	23
2.5.2.	Ciclo de vida.....	24
2.5.2.1.	Planificación del proyecto	25
2.5.2.2.	Definición de requerimiento del negocio	25
2.5.2.3.	Modelado dimensional	26
2.5.2.4.	Diseño físico	27
2.5.2.5.	Diseño e implementación del subsistema de Extracción, Transformación, Carga (ETL)	27
2.5.2.6.	Implementación.....	28
2.5.2.7.	Mantenimiento y crecimiento de un data warehouse	28
2.5.2.8.	Especificaciones de aplicaciones de BI	28
2.5.2.9.	Diseño de la arquitectura técnica.....	28
2.5.2.10.	Selección de productos e implementación	29
3.	METODOLOGÍA.....	30
3.1.	Planificación y administración del proyecto	30
3.1.1.	Definición de requisitos del proyecto.....	30
3.1.2.	Modelo dimensional	31
3.1.3.	Modelo físico	31
3.1.4.	Diseño e implementación del proceso ETL	32
3.1.5.	Diseño de la arquitectura técnica	33
3.1.6.	Selección del producto	33
3.1.7.	Aplicaciones de BI.....	33
4.	RESULTADOS	34
4.1.	PLANIFICACIÓN Y ADMINISTRACIÓN DEL PROYECTO	34
4.1.1.	Definición requisitos del proyecto.....	34
4.1.2.	Modelo dimensional	35
4.1.2.1.	Nivel de granularidad	35
4.1.2.2.	Identificar las dimensiones y mapeo de datos	36
4.1.3.	Modelo físico	39
4.1.4.	Diseño e implementación del proceso ETL	40
4.1.4.1.	Extracción	40

4.1.4.2.	Transformación	41
4.1.4.3.	Carga	43
4.1.5.	Diseño de la arquitectura técnica	43
4.1.5.1.	Back Room	43
4.1.5.2.	Front Room	44
4.1.6.	Selección del producto	44
4.1.7.	Aplicaciones de BI	45
4.1.8.	Implementación	45
4.1.9.	Mantenimiento y crecimiento	46
4.2.	MODELO DE GESTIÓN DE CONOCIMIENTO DE LA UNIVERSIDAD	
	47	
4.2.1.	Departamento de Biblioteca	47
4.2.2.	Instituto de transferencia de tecnologías	48
4.2.3.	Instituto de informática y computación	49
4.2.4.	Modelo de conocimiento del área de planificación	49
4.3.	DESARROLLO DEL DATA WAREHOUSE	50
4.3.1.	Desarrollo de los data mart	50
4.3.1.1.	Características de los data mart	51
4.3.1.2.	Implantación de los data mart	51
4.3.1.3.	Pruebas de funcionamiento en los data mart	52
4.3.1.3.1.	Generación de cubos OLAP	54
4.4.	EVALUACIÓN Y DESPLIEGE	59
5.	CONCLUSIONES Y RECOMENDACIONES	70
5.1.	CONCLUSIONES	70
5.2.	RECOMENDACIONES	71
	BIBLIOGRAFÍA	72

ÍNDICE DE TABLAS

	PÁGINA
Tabla 1. Fuentes de generación del conocimiento (Zapata, 2004).....	9
Tabla 2. Enfoques de la minería de datos educacionales (García, 2010) ...	23
Tabla 3. Esquema base de origen.....	35
Tabla 4. Mapeo de datos del data mart de planificación institucional.....	36
Tabla 5. Mapeo de datos del data mart de acreditación institucional	37
Tabla 6. Evaluación de productos	44
Tabla 7. Tamaño de disco duro	45
Tabla 8. Total de filas procesadas en los data marts	46
Tabla 9. Variables del departamento de biblioteca.....	48
Tabla 10. Variables del departamento de Investigación y Transferencia de Tecnologías	48

ÍNDICE DE FIGURAS

PÁGINA

Figura 1. Generación de información	5
Figura 2. Generación de conocimiento	6
Figura 3. Etapas del proceso KDD (Ril et al., 2013).....	10
Figura 4. Modelo Nonaka y Takeuchi (Sánchez, 2005)	11
Figura 5. Modelo Technology Broker (Gonzalez, 2009).....	13
Figura 6. Arquitectura de inteligencia de negocios (Vélez, 2011)	14
Figura 7. Arquitectura del data warehouse (Connolly & Begg, 2005).....	18
Figura 8. Data Warehouse (Espinosa, 2010)	24
Figura 9. Ciclo de vida Kimball (Rivadera, 2010)	25
Figura 10. Diagrama del flujo de proceso del proceso dimensional Kimball (Rivadera, 2010)	27
Figura 11. Matriz de evaluación de productos (Ilbay, 2009).....	29
Figura 12. Información necesaria en el área de planificación	34
Figura 13. Origen de datos	40
Figura 14. Asignación de columnas de entrada y salida	40
Figura 15. Transformación de valores nulos	41
Figura 16. Transformación de cadenas extensas	42
Figura 17. Transformación de conversión de datos	42
Figura 18. Editor de destino de SQL	43
Figura 19. Complemento de minería de datos en Microsoft Excel	45
Figura 20. Propuesta de modelo de gestión de conocimiento para planificación UTE	49
Figura 21. Flujo de datos en dimensiones	52
Figura 22. Flujo de datos en tablas de hechos	52
Figura 23. Error de campo “NULL” localizados en los datos de carrera.....	53
Figura 24. Error de escritura en las base de datos de origen	53
Figura 25. Fechas de nacimiento de docentes con valores nulos.....	54
Figura 26. Pantalla de creación de un proyecto multidimensional	54
Figura 27. Pantalla de selección de nuevos orígenes de datos	55
Figura 28. Pantalla de una nueva conexión	55
Figura 29. Pantalla de nuevas vistas de origen de datos.....	56
Figura 30. Pantalla de selección de tablas para crear la vista	56
Figura 31. Pantalla de un nuevo cubo.....	57
Figura 32. Pantalla del asistente para cubos grupo de medidas.....	57
Figura 33. Pantalla asistente para cubos dimensiones	58
Figura 34. Pantalla de procesamiento del cubo	58
Figura 35. Pantalla de Microsoft Excel de conexión al cubo	59
Figura 36. Pantalla informes de planificación por códigos KPI y por nombres de unidad	60
Figura 37. Pantalla informe de planificación por códigos KPI con gráficos .	60
Figura 38. Evaluación de metas del año 2013 y 2014 con un cumplimiento de 0 y 100%.....	61
Figura 39. Estudiantes graduados por año, por sexo, por facultad	61
Figura 40. Estudiantes matriculados por año	62
Figura 41. Estudiantes matriculados por modalidad	62

Figura 42. Configuración del servidor SQL para minería de datos con Office	63
Figura 43. Pantalla del Add-in de minería de datos en Microsoft Excel	64
Figura 44. Asistente para crear el modelo y clasificar	64
Figura 45. Parámetros de los algoritmos de clasificación	65
Figura 46. División de los datos de entrenamiento y conjunto de pruebas .	65
Figura 47. Pantalla árbol de decisión en estado de graduación.....	66
Figura 48. Asistente para agregar un modelo a la estructura	66
Figura 49. Pantalla de perfiles de atributos con Bayes Naive de Microsoft.	67
Figura 50. Gráfico de precisión de los algoritmos	68
Figura 51. Pantalla de árboles de decisión de los estudiantes de Ingeniería Informática	68
Figura 52. Pantalla de perfiles del atributo de los estudiantes de Ingeniería Informática	69
Figura 53. Gráfico de precisión del modelo de estudiantes de Ingeniería Informática	69

ÍNDICE DE ANEXOS

	PÁGINA
Anexo 1. Tabla de hechos porcentaje meta.....	78
Anexo 2. Dimensión inversión.....	79
Anexo 3. Dimensión código KPI.....	80
Anexo 4. Dimensión unidades.....	80
Anexo 5. Dimensión objetivo unidad.....	80
Anexo 6. Dimensión tiempo.....	81
Anexo 7. Dimensión código meta mes.....	81
Anexo 8. Dimensión meta cumplirse.....	82
Anexo 9. Dimensión meta indicador logro.....	82
Anexo 10. Dimensión inversión gasto.....	82
Anexo 11. Dimensión logros alcanzados.....	83
Anexo 12. Dimensión objetivo institucional.....	83
Anexo 13. Dimensión objetivo unidad.....	83
Anexo 14. Dimensión meta plan de mejora.....	84
Anexo 15. Dimensión meta retroalimentación.....	84
Anexo 16. Dimensión meta usuario unidad.....	84
Anexo 17. Dimensión biblioteca.....	85
Anexo 18. Dimensión campus.....	85
Anexo 19. Dimensión aulas.....	85
Anexo 20. Dimensión presupuesto.....	86
Anexo 21. Dimensión materias.....	86
Anexo 22. Dimensión carreras.....	86
Anexo 23. Dimensión facultad.....	86
Anexo 24. Dimensión docente.....	87
Anexo 25. Dimensión tiempo dedicación.....	88
Anexo 26. Dimensión período sabático.....	88
Anexo 27. Dimensión ingreso concurso.....	88
Anexo 28. Dimensión sub área último título cuarto nivel.....	89
Anexo 29. Dimensión período.....	89
Anexo 30. Dimensión cursando estudios.....	89
Anexo 31. Dimensión cargo autoridad.....	90
Anexo 32. Dimensión último título cuarto nivel.....	90
Anexo 33. Dimensión categoría personal académico.....	91
Anexo 34. Dimensión último título tercer nivel.....	91
Anexo 35. Dimensión sexo.....	91
Anexo 36. Dimensión modalidad.....	92
Anexo 37. Dimensión colegio.....	92
Anexo 38. Dimensión discapacidad.....	92
Anexo 39. Dimensión estado civil.....	92
Anexo 40. Dimensión localidad.....	93
Anexo 41. Dimensión estudiantes primer nivel.....	93
Anexo 42. Dimensión estudiante posgrado.....	94
Anexo 43. Dimensión publicación artículos.....	95
Anexo 44. Dimensión libro capítulo.....	95

Anexo 45. Dimensión estudiantes pregrado	96
Anexo 46. Dimensión revista indexada	97
Anexo 47. Dimensión proyectos de vinculación	97
Anexo 48. Dimensión proyectos de investigación	98
Anexo 49. Tabla de hechos estudiantes primer nivel	99
Anexo 50. Tabla de hechos internet.....	100
Anexo 51. Tabla de hechos proyectos de vinculación	100
Anexo 52. Tabla de hechos presupuesto universitario.....	101
Anexo 53. Tabla de hechos estudiantes posgrado	102
Anexo 54. Tablas de hechos biblioteca.....	102
Anexo 55. Tabla de hechos aulas	103
Anexo 56. Tablas de hechos libros artículos.....	103
Anexo 57. Tabla de hechos periodo sabático	104
Anexo 58. Tabla de hechos docentes	105
Anexo 59. Tabla de hechos estudiantes pregrado	106
Anexo 60. Tabla de hechos proyectos de investigación	107
Anexo 61. Tabla de hechos asignatura docentes pregrado	107
Anexo 62. Tabla de hechos cursando estudios	108

NOMEMCLATURA

UTE: Universidad Tecnológica Equinoccial.

TIC: Tecnologías de la Información y la Comunicación.

TI: Tecnologías de Información.

ETL: *Extract, Transform and Load* - Extracción, Transformación y Carga.

ITIC – UTE: Instituto de Tecnologías de Informática y Computación UTE.

SABUTE: Sistema de Automatización de Biblioteca UTE

SICAF - UTE: Sistema Integrado Contable Académico y Financiero UTE.

CEAACES: Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior.

KPI: *Key Performance Indicators* - Indicadores Claves de Rendimiento.

KDD: *Knowledge Discovery in Databases* - Descubrimiento de Conocimiento en Base de Datos.

ERP: *Enterprise Resource Planning* - Software de Gestión Integrada.

CRM: *Customer Relationship Management* - Software para la Administración de la Relación con los Clientes.

SQL: *Structured Query Language* – Lenguaje Estructurado de Consultas.

OLTP: *OnLine Transaction Processing* - Procesamiento de Transacciones en Línea.

OLAP: *On-Line Analytical Processing* - Procesamiento Analítico en Línea

BD: Base de datos

BI: *Business Intelligence* – Inteligencia de Negocios

GHz: Gigahercio

RESUMEN

La información se ha convertido en un elemento de gran importancia para el desarrollo de las instituciones por lo cual es importante conocer el ambiente en el que se desenvuelven para tomar acertadas decisiones. Por esta razón los datos históricos y las bases de datos operacionales depuradas y unificadas son fuente de información valiosa para cada institución.

El proyecto de titulación está orientado a proporcionar una herramienta informática que ayude con la toma de decisiones en el departamento de planificación. Este departamento es encargado de establecer los diferentes planes para el proceso de acreditación institucional, proceso que es realizado cada cinco años por el CEAACES y se basa en la eficiencia académica, investigación, organización e infraestructura.

Tomando en cuenta la problemática mencionada anteriormente, para el presente proyecto se desarrolla un *data warehouse* con herramientas de inteligencia de negocios y minería de datos para facilitar la toma de decisiones oportunas ayudando a identificar debilidades en los procesos de la institución y rectificarlos de una manera eficiente, convirtiéndose en fortalezas para la universidad. Para el proceso de desarrollo del *data warehouse* se estableció como metodología la de *Ralph Kimball* permitiendo una planificación desde el inicio hasta el final de proyecto. Esta propuesta incluye un modelo de gestión del conocimiento y los diferentes requisitos que necesita el departamento para que el *data warehouse* pueda tener un mantenimiento y un crecimiento posterior.

Para el aprovechamiento de esta herramienta se utilizó *Microsoft Excel* con complementos de minería de datos que facilitan la creación de reportes dinámicos y el análisis de modelos predictivos en busca de tendencias desconocidas que apoyen al proceso académico e institucional.

ABSTRACT

The use of information in time has become a relevant element in the development of enterprises and institutions. Information helps people to recognize the internal and external advantages and disadvantages and to understand the environment in which they operate. The main objective is to visualize mistakes, evaluate and make good decisions that will improve the performance of organizations. For this reason, the use of historical facts and operational databases are valuable sources of knowledge.

The objective of this academic project is to implement an informatic tool that supports the decision making process in the planning department of the university. This department is in charge of developing strategies and plans that contribute to the institutional process of accreditation by CEAACES. This process is done every five years, and is based on academic efficiency, research development, and the correct organization and use of infrastructure.

Taking into account this problem, in this project a data warehouse with business intelligence and data mining tools was developed to facilitate proper decision making, to identify weaknesses in institutional processes and to propose solutions that are viable and appropriate to strengthen the university. For the development process of the data warehouse a knowledge management model was used which takes into account the requirements established in Ralph Kimball methodology. This methodology supports the planning process from the beginning to the end of the project, which is reflected in a proposal that takes into account the requirements of the planning department for the proper operation of the tool and its maintenance over time.

Data mining complements of Microsoft Excel are used for the efficient job and exploitation of the data warehouse. This facilitate the creation of dynamic reports and analysis of predictive models and the discovery of unknown trends that contributes to the academic and the institutional process.

1. INTRODUCCIÓN

1. INTRODUCCIÓN

La Universidad Tecnológica Equinoccial (UTE) forma parte del sistema de Educación Superior que cumple con el propósito de formar a nuevos profesionales en actividades con proyección futura mediante la investigación y el planteamiento de soluciones para los problemas del país.

La UTE cuenta con el departamento de planificación, encargado de dar un seguimiento permanente de cada área, mediante metas y objetivos que pretende alcanzar la institución, contribuyendo principalmente a la mejora continua mediante estrategias institucionales que fomenten el proceso de aprendizaje y el desarrollo organizacional. El departamento además de mantener una planificación estratégica de desarrollo institucional, se responsabiliza de la evaluación y acreditación institucional que es llevada a cabo por los organismos de control gubernamentales como es El Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES).

La UTE comprometida con su misión busca siempre mejorar la calidad de la educación para alcanzar altos estándares de desempeño que faciliten la evaluación por parte del CEAACES perfeccionando cada uno de los factores de desempeño de evaluación en la acreditación institucional como son la academia, investigación, eficiencia académica, infraestructura, organización. Por tal motivo la información se vuelve una de las piezas fundamentales para la institución debido a que ayuda a la toma de decisiones. Además contribuye como herramienta importante al crecimiento institucional mediante el cumplimiento de objetivos y metas de la organización.

Hoy en día muchas organizaciones tienen información en grandes cantidades pero no saben cómo gestionarla de una manera eficiente y que puedan ayudar a la toma de decisiones de negocio. Los avances de la tecnología y el acceso cada vez mayor a internet han facilitado un mayor acceso a la información. Esta información se encuentra dispersa en diferentes fuentes y es necesario

realizar un análisis exhaustivo del negocio que facilite recopilar y administrar esta información y esquematízala para que los usuarios interesados en la organización puedan hacer uso de la misma y obtener ventaja competitiva.

La inteligencia de negocios es un término que se empezó a utilizar a finales de los años 90 y es un componente de alto nivel de los sistemas de información que tiene una serie de ventajas y beneficios para las organizaciones. Uno de los mayores beneficios es que permiten analizar y depurar la información de las bases de datos transaccionales para transformarla y convertirla en conocimiento. Este conocimiento soporta a los directivos para tomar decisiones estratégicas, mejorando el desempeño de muchas funciones organizacionales: *marketing*, ventas, compras, estadística, bodega, finanzas, atención al cliente, postventa, etc.

Las herramientas de inteligencia de negocios facilitan la visualización de la información en las organizaciones a partir del uso de reportes y *dashboards* que transforman simples datos en conocimiento con beneficio para la institución aumentando el rendimiento en los negocios.

Para el planteamiento del problema se analizó los diferentes procesos que lleva a cabo el área de planificación de la UTE para obtener información de las estrategias institucionales de cada unidad, con sus metas y objetivos de acuerdo a los diferentes periodos académicos. La información manejada por el departamento para la acreditación institucional es reunida mediante solicitudes a cada unidad encargada de llenar sus respectivas variables en un formato predefinido en *Microsoft Excel*.

La problemática radica en la dificultad de tener una información precisa y amplia debido al tiempo que se toma en obtener datos de cada unidad mediante solicitudes, lo cual dificulta tener una información útil y necesaria en el momento oportuno, dificultando realizar un análisis profundo que ayude en la toma de decisiones. Tampoco se cuenta con un modelo de gestión del conocimiento para administrar de mejor manera los recursos y la información

que se genera tanto en el área académica como administrativa. Más aún son escasas las herramientas de TI que apoyan la creación, uso y transferencia de conocimiento en la universidad.

Para solucionar este problema se propuso una solución basada en TI mediante la implementación de un *data warehouse* para obtener datos integrados, consistentes, no volátiles y confiables poniendo a disposición de los usuarios autorizados información pertinente y oportuna.

La creación de una propuesta de un modelo de gestión del conocimiento ayudará en la administración de los procesos principales para aumentar la capacidad de generar y recoger información en el departamento.

Las herramientas de inteligencia de negocios ayudarán en el proyecto para mejorar el control y el seguimiento continuo de algunos procesos que maneja el área de planificación, lo que será determinante para una mejora en la toma de decisiones por parte de los usuarios y directivos. Esto a la vez encaminará a la consecución de las metas y objetivos programados.

La minería de datos contribuirá en identificar relaciones, patrones, tendencias desconocidas o difíciles de identificar a simple vista, con el fin de tener una mejor perspectiva en la generación de nuevas estrategias competitivas para fortalecer el mejoramiento continuo de la universidad.

De acuerdo a lo planteado anteriormente se propone para el proyecto como objetivo general:

Crear para el área de planificación de la Universidad Tecnológica Equinoccial (UTE) un *data warehouse* que mediante un modelo de gestión del conocimiento y aplicando técnicas de minería de datos educacionales permita mejorar la toma de decisiones por partes de las autoridades.

Conforme al objetivo general estipulado anteriormente se postula los siguientes objetivos específicos.

Utilizar la metodología de *Ralph Kimball* como guía para el desarrollo del proyecto.

Diseñar un modelo de gestión del conocimiento para el área de planificación de la UTE.

Determinar los principales requerimientos que el área de planificación de la UTE busca en una herramienta de Inteligencia de Negocios.

Diseñar reportes que permitan gestionar la información requerida e implementar técnicas de minería de datos para descubrir tendencias para mejorar alguno de los indicadores valiosos para la UTE.

2. MARCO TEÓRICO

2. MARCO TEÓRICO

2.1. GESTIÓN DEL CONOCIMIENTO

2.1.1. Dato

Los datos son letras, números que describen situaciones o sucesos relacionado a personas o cosas.

“Los datos desempeñan un papel similar al de las materias primas en el proceso de producción de bienes industriales, y necesitan ser manipulados para que generen valor” (Velasco & Quintana, 2003).

2.1.2. Información

La información es el conjunto de datos transformados o procesados que generan relevancia y utilidad de algo nuevo que se desconocía (Casanova, 2009).

La Figura 1 indica el proceso de generación de información.

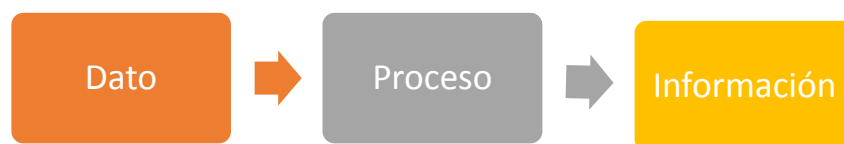


Figura 1. Generación de información

2.1.3. Conocimiento

El conocimiento está vinculado exclusivamente con la facultad de aprendizaje, entendimiento de los seres humanos para actuar o tomar decisiones inteligentemente a partir de nuevas fuentes de información desarrollando la capacidad de adquirir nuevos conocimientos.

“La gran diferencia entre los procesos de generación de información y generación de conocimiento es que en este último caso es necesaria la intervención de un ser humano” (Casanova, 2009).

La Figura 2 indica el proceso de generación de conocimiento.

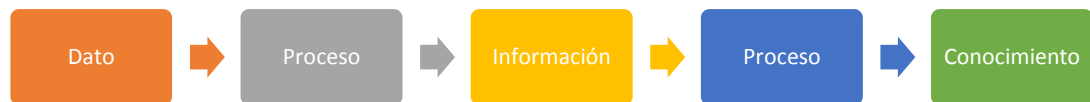


Figura 2. Generación de conocimiento

2.1.4. Conocimiento explícito

“El conocimiento explícito es aquel que puede ser capturado fácilmente, codificado, almacenado y transmitido en un lenguaje formal y sistemático.” (Boronat, Villar, & Puig, 2007). Este conocimiento es expresado mediante palabras o números resultando fácil compartirlo en forma de datos como: formulas científicas, especificaciones de productos, productos de TI. En las organizaciones es más evidente mediante los procedimientos de trabajo como son las estrategias de cada empresa.

2.1.5. Conocimiento tácito

El conocimiento tácito o implícito formado por las destrezas, capacidades mentales, habilidades técnicas llamadas *know-how*, en la que cada uno de nosotros tenemos arraigadas en nuestro interior y que no puede ser directamente compartido, es conveniente aprovecharlo en actividades productivas. (Boronat et al., 2007) (Contreras, 2010).

2.1.6. Gestión del conocimiento

La gestión del conocimiento persigue la creación de ventajas competitivas sostenibles mediante aprendizajes organizacionales y continuos con la ayuda de las TIC, favoreciendo en la recopilación y el procesamiento de información mediante plataformas digitales (Benavides & Quintana, 2005).

“Las tecnologías de la información no pueden ser programadas para detectar los cambios dinámicos en el entorno empresarial y tienen dificultad para recoger el conocimiento de tipo tácito” (Velasco & Quintana, 2003), el cual es usado para dar sentido a la información.

Las ventajas competitivas actualmente no solo están dadas por la información que se puede recopilar en las empresas sino por el conocimiento que se pueda adquirir para el desempeño de la organización.

2.1.7. Necesidades de información

Las necesidades de información surgen de la gran cantidad de información escrita, sonora, visual el cual el ser humano necesita procesar para generar un nuevo conocimiento y poder adaptarse al medio ambiente. Como es el caso de las organizaciones que son grandes generadoras de información y necesitan adaptarse rápidamente a los nuevos entornos en los que se encuentran las instituciones manejadas por personas que son los proveedores de las diferentes ideas (Calva, 2013).

La información se ha convertido para las empresas u organizaciones en uno de los activos más importantes y se comienza a tratar a los datos de una manera relacionada para poder ayudar a la toma de decisiones de una manera más metodológica (Nader, 2011).

2.1.8. Fuentes de generación de conocimiento

La generación de conocimiento es una herramienta fundamental en las empresas exitosas debido a que permite mejorar y aumentar su rendimiento competitivo.

El proceso de generación se realiza dentro de la organización el cual comprende de tres actividades: (Zapata, 2004).

1. La creación interna de conocimiento.
2. El aprendizaje por acción.
3. La adquisición y el acceso del conocimiento externo.

1. **La creación interna de conocimiento.** Es un proceso en el cual la organización incrementa el conocimiento mediante el trabajo conjunto de las personas y lo concreta como parte del sistema de aprendizaje de la empresa (Zapata, 2004).

“Una organización no puede crear conocimientos sin las personas que la integran, lo que hacen las empresas es proporcionar el entorno adecuado” (Zapata, 2004).

El conocimiento también puede ser generado informalmente mediante reuniones de diversos intereses, correo electrónico, grupos de trabajos virtuales o telefónicamente, en el cual se comparte el conocimiento para la solución de problemas y poder colaborar con la eficacia de la organización.

El conocimiento generado dentro de la empresa es especialmente valioso ya que tiende a ser único, específico y con un gran componente tácito, haciéndolo más difícil de imitar por los competidores (Zapata, 2004).

2. **Aprendizaje por acción.** Este proceso implica entender cómo funcionaba anteriormente la organización, y como fue incorporada en las rutinas organizativas que permiten convertir el conocimiento de las personas en un conocimiento colectivo (Zapata, 2004).

Las organizaciones aprenden de dos formas:

- Aprendizaje de sus miembros.
- Aprendizaje de los nuevos miembros quienes no cuentan con un conocimiento previo de la organización.

“Las empresas construyen una base de conocimientos mediante las experiencias que sus empleados han aprendido y aplicado dentro del marco estratégico de la organización” (Zapata, 2004).

3. **Adquisición de conocimiento externo.** Este proceso implica compartir conocimientos tácitos con otras organizaciones mediante alianzas suponiendo una oportunidad directa e indirecta de tener acceso a las habilidades, tecnologías, mercados, competencias

esenciales e incluso información estratégica con sus socios o proveedores.

Muy pocas empresas tienen un amplio conocimiento de todo lo requerido para un desarrollo tecnológico continuo, por lo que se recurre a fuentes externas, reduciendo el tiempo de espera necesario para el desarrollo interno del conocimiento (Lidia, Aguiar, & Saá, 2008). Los clientes son una fuente valiosa de información, para el desarrollo o mejora de un nuevo producto o servicio disponible en el mercado.

La Tabla 1 muestra una síntesis de las distintas formas de generación de conocimiento.

Tabla 1. Fuentes de generación del conocimiento

(Zapata, 2004)

CREACIÓN DE CONOCIMIENTO	APRENDIZAJE POR ACCIÓN	ADQUISICIÓN Y ACCESO DE CONOCIMIENTO EXTERNO
Investigación	Entrenamiento en el trabajo.	Alianzas
Diseño	Experimentos	Proveedores
Socialización	Simulaciones	Clientes
Externalización		Universidades
Combinación		Agencias gubernamentales
Internalización		Implementación de sistemas de información

2.1.8.1. Descubrimiento del conocimiento en bases de datos

El descubrimiento de conocimiento en las bases de datos conocido en inglés como *Knowledge Discovery in Databases* (KDD), se relaciona con el proceso de encontrar o descubrir patrones potencialmente útiles dentro de algún repositorio de datos (Silva, 2007).

Las etapas del proceso de KDD son los siguientes: (Ril, Rodríguez, Piñero, & Martínez, 2013).

- **Selección de datos.**- Se escoge los repositorios o fuentes de datos.
- **Preprocesado de datos.**- En esta fase se limpian los datos que se extrae desde los orígenes.
- **Transformación de datos.**- Consiste en transformar y generar nuevas variables a partir de las ya existentes.
- **Minería de datos.**- En esta etapa se aplican diferentes algoritmos de minería de datos con el objetivo de extraer patrones previamente desconocidos.
- **Interpretación.**- Se interpreta los resultados obtenidos de acuerdo a los métodos utilizados en la etapa anterior.

En la Figura 3 se observa el proceso de descubrimiento de conocimiento en las bases de datos.



Figura 3. Etapas del proceso KDD

(Ril et al., 2013)

2.1.9. Modelos para la gestión del conocimiento

2.1.9.1. Modelo de creación de conocimiento organizacional Nonaka y Takeuchi

Este modelo distingue entre el conocimiento tácito y el explícito para la creación del conocimiento organizacional, se desarrollan un modelo de cuatro etapas que comprende: (Avilés, 2014).

- Socialización.
- Externalización.
- Combinación.
- Interiorización.

La Figura 4 muestra el modelo de conocimiento organizacional según Nonaka y Takeuchi.

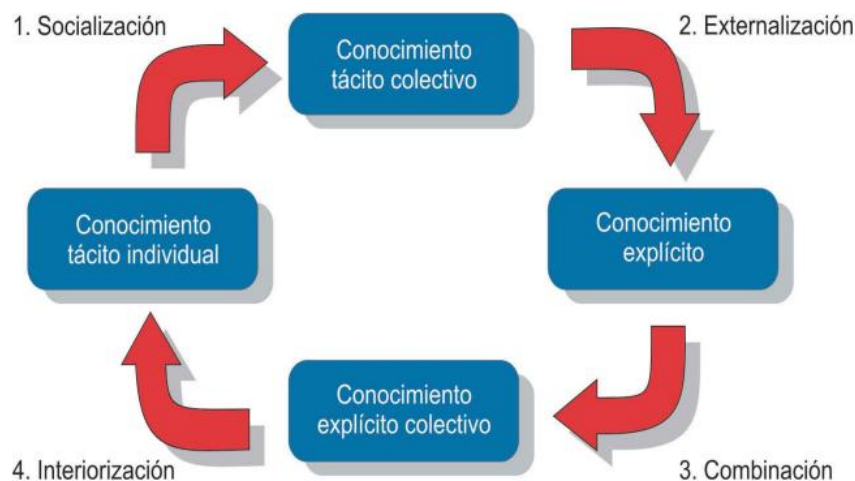


Figura 4. Modelo Nonaka y Takeuchi

(Sánchez, 2005)

1. **Socialización.** Este proceso se evidencia claramente cuando se realiza capacitaciones, donde el instructor comparte sus experiencias, habilidades, sus aprendices adquieren estos nuevos conocimientos mediante la observación, imitación y la práctica convirtiéndose en un conocimiento tácito (Sánchez, 2005).
2. **Exteriorización.** Esta fase se convierte el conocimiento tácito en explícito, utilizando el dialogo para formar conceptos explícitos como hipótesis o modelos (Pávez Salazar, 2000).
3. **Combinación.** Esta etapa se evidencia cuando se intercambia un conocimiento sintetizado de interés para la institución, por medio de informes, presentaciones, convirtiéndose en un conocimiento explícito al momento que este conocimiento sintetizado llega a las personas de la institución (León & Mercader, 2002).

4. **Internalización.** Esta parte del proceso convierte el conocimiento explícito a tácito formado mediante la socialización de las dos etapas anteriores, se forma modelos mentales, diagramas, manuales que faciliten la transferencia del conocimiento explícito a otras personas convirtiéndose en activos importantes para la organización (Sánchez, 2005).

2.1.9.2. Modelo de creación de conocimiento *Technology Broker*

El modelo de *Broker* busca conocer las capacidades de la organización para lograr sus metas al poder planear una investigación para su desarrollo. *Broker* se sustenta en que la suma del valor intangible y el capital para aumentar el valor de la organización en el mercado (Langa, Carbonell, Vicedo, & Miquel, 2005).

Los activos intangibles de acuerdo al modelo se clasifican:

1. **Los activos del mercado.** Se otorga una ventaja competitiva como a las marcas, clientes, franquicias. Esta categoría denota una identidad única de los productos y servicios que se ofrecen (Hernández, 2003).
2. **Los activos humanos.** Se hace referencia a los niveles académicos, los conocimientos y habilidades que no son propiedad de la empresa.
3. **Los activos de propiedad intelectual.** En esta etapa se incluye el valor adicional que da la organización como son: las patentes, derechos de autoría, derechos de diseño (Gonzalez Millan, 2009).
4. **Los activos de infraestructura.** Son aquellos procesos tecnológicos o métodos que permiten el funcionamiento de la institución como son: sistemas de información, cultura organizativa, bases de datos existentes (Funes & Hernández, 2001).

La Figura 5 muestra las cuatro categorías del modelo *Technology Broker*, iniciando desde los objetivos corporativos y el capital intelectual de toda la organización.

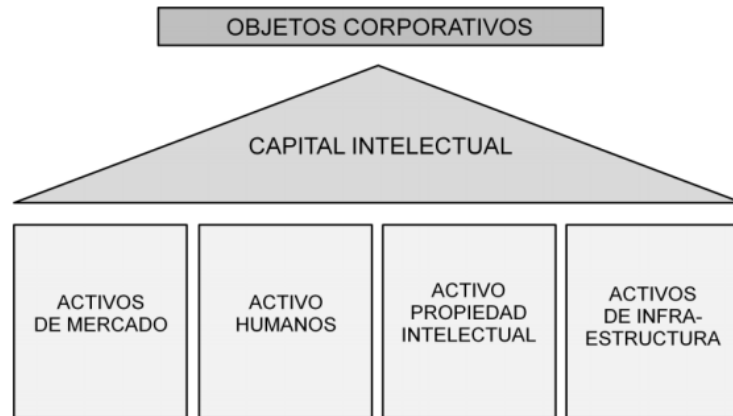


Figura 5. Modelo *Technology Broker*

(Gonzalez , 2009)

2.2. INTELIGENCIA DE NEGOCIOS

2.2.1. Definición de inteligencia de negocios

Según *Data Warehouse Institute*, lo define como la combinación de tecnologías, herramientas y procesos que permiten transformar los datos almacenados en conocimiento dirigido a un plan o estrategia comercial (Oracle, 2009).

La inteligencia de negocios utiliza metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos de una manera estructurada para una determinada área, generando conocimiento para que puedan ser distribuidos a las personas que lo necesitan en el momento oportuno a fin de que sirva como apoyo en la toma de decisiones (Gomez, 2010).

2.2.2. Arquitectura de inteligencia de negocios

Una solución de inteligencia de negocios comienza en las fuentes de datos usadas en la institución como archivos de *Excel*, *ERP*, *CRM*, preliminarmente es necesario aplicar transformaciones de depuración para tener un proceso estructurado de datos.

La Figura 6 indica una arquitectura de inteligencia de negocios basados en herramientas de *Microsoft*.

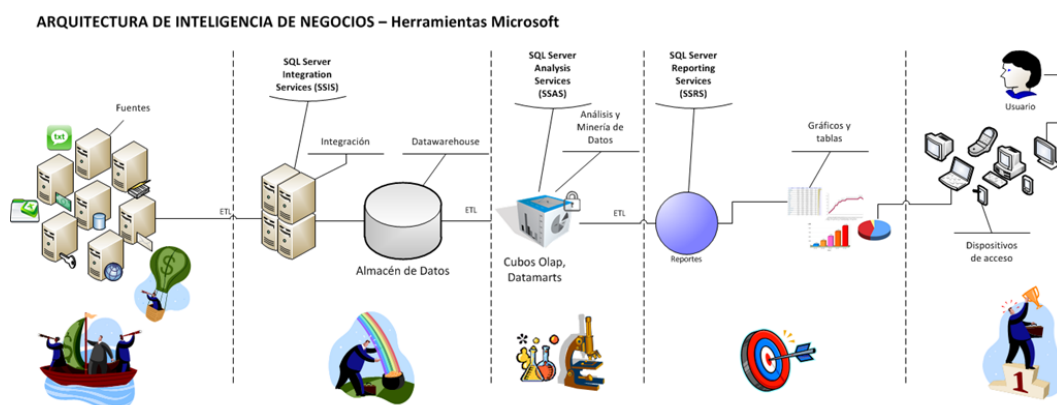


Figura 6. Arquitectura de inteligencia de negocios

(Vélez, 2011)

2.2.2.1. Fuentes de datos

Es la base de datos relacional, hojas de cálculo, archivos planos, que sirven de almacenamiento de la información, la cual es generada por los usuarios diariamente en la institución u organización.

2.2.2.2. Integración de datos

En esta etapa se integran todos los datos definidos previamente en el análisis y diseño de la solución de inteligencia de negocios que serán enviados al *data warehouse* construido (Vélez, 2011).

2.2.2.3. Análisis de datos

En esta fase se tiene un *data warehouse* poblado y estructurado completamente, se hacen uso de las herramientas de análisis para poder generar cubos multidimensionales. Estos se reestructuran dependiendo de los requerimientos de la organización, se crean jerarquías, se oculta información innecesaria, se realiza minería de datos para obtener un análisis completo de los datos verificando que los mismos se están cargando consistentemente. (Vélez, 2011).

2.2.2.4. Reportes

En esta etapa se utiliza herramientas de visualización para el usuario como *SQL Server Reporting Services* que facilita el diseño de reportes y *dashboards* corporativos ajustándose a las necesidades del cliente final mediante la información generada anteriormente.

2.2.2.5. Usuario

Son las personas que tendrán acceso a los reportes y *dashboards* mediante diferentes herramientas para ayudar a tomar decisiones dentro de la institución (Vélez, 2011).

2.2.3. Componentes

2.2.3.1. Diseño conceptual de los sistemas

Para el diseño conceptual de un modelo de inteligencia de negocios se debe resolver tres preguntas básicas como: ¿Cuál es la información requerida para la gestión en el departamento?; ¿Cuál debe ser el formato a utilizarse?, ¿Cuál es la composición de los datos a utilizar? (Rotaeché, 2007).

En la primera fase se construye el *data warehouse* y *data marts* estructurando la información con los diferentes niveles de criterios. En la segunda fase se desarrolla el análisis de criterios de los directivos mediante sus respectivas necesidades de información (Rotaeché, 2007).

2.2.3.2. Construcción y alimentación del data warehouse

Un *data warehouse* es una réplica de los datos transaccionales debidamente estructurados, depurados y seleccionados para las diferentes actividades de consultas y reportes. (Aguirre, Andrade, Maldonado, & Ureta, 2006). “La vocación del *data warehouse* es separar los sistemas operacionales de las necesidades de información para la gestión, de forma que cambios en aquéllos no afecten a éstas, y viceversa” (Rotaeché, 2007).

2.2.3.3. Herramientas de explotación del data warehouse

2.2.3.3.1. Consultas y reportes

Herramientas para la construcción de informes en detalle de acuerdo a las necesidades de la institución a partir de la información guardada en el *data warehouse*.

2.2.3.3.2. Cuadro de mando analítico

Herramientas para diseño de informes pequeños con KPI importantes para la institución que permiten una análisis rápido de los resultados para los usuarios (Rotaeché, 2007).

2.2.3.3.3. Cuadro de mando integral

Herramientas para la construcción de cuadros de mando integral o *dashboards* que proporcionan información constante del cumplimiento de los indicadores importantes para la institución (Kaplan & Norton, 2005).

2.3. DATA WAREHOUSE

2.3.1. Definición de data warehouse

Es un almacén o repositorio de datos orientado a una determinada empresa o institución, para apoyar las decisiones gerenciales. La información almacenada en un *data warehouse* está integrada de acuerdo a la orientación del tema, es no volátil y variable con el tiempo (Guanotásig, 2013).

Un *data warehouse* es una simple percepción, debido a que llega a ser una base de datos con información clave manejada por la institución, que al integrarla y procesarla podemos obtener una perspectiva más amplia de la realidad corporativa (Valencia Arcos & Guevara Lenis, 2007). La información almacenada es de fácil accesibilidad para los usuarios debido a que tiene un registro histórico organizado para poder ser utilizados dentro de una perspectiva de tiempo.

2.3.2. Características un data warehouse

2.3.2.1. Orientado al tema

La información se organiza de acuerdo a los aspectos de interés para la empresa centrándose en los departamentos claves que son necesarios para el análisis de información. La información almacenada está en contraste con la generada en los ambientes operaciones de la institución como por ejemplo estudiantes, docentes, asignaturas (Rojas, 2009).

2.3.2.2. Integración

La información almacenada debe estar integrada dando a cada tipo de datos un estándar de una manera general, debido a que normalmente los datos operacionales tienen errores de inconsistencia, uniformidad y diferente codificación (Duque, 2010).

2.3.2.3. De tiempo variante

La información del *data warehouse* es requerida en algún momento es decir no “ahora mismo”, esta característica es diferente al de los sistemas de ambiente operacional, en el cual se necesita la información al instante para acceder al trabajo regular. La información histórica no es muy importante en los ambientes operacionales, el *data warehouse* es muy importante para poder evaluar estadísticas y diferentes tendencias (Rojas, 2009).

2.3.2.4. No volátil

La información almacenada es estable debido a que un *data warehouse* no sufre operaciones como inserción (*insert*), eliminación (*delete*), modificación (*update*), que comúnmente se utiliza en los sistemas operacionales. Un *data warehouse* realiza dos operaciones que son la carga de datos y el acceso a los mismo (Duque, 2010).

2.3.3. Arquitectura de un data warehouse

La arquitectura de un *data warehouse* es una forma para la representación de la estructura total de los datos en la institución como lo muestra la Figura 7.

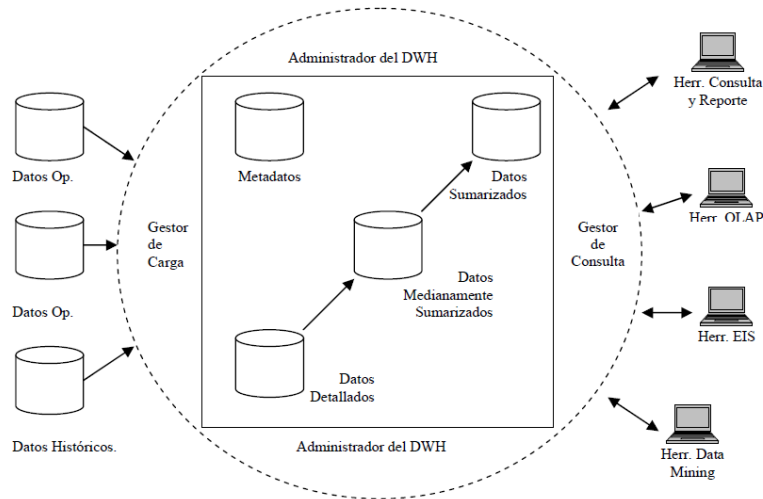


Figura 7. Arquitectura del *data warehouse*

(Connolly & Begg, 2005)

2.3.3.1. Componentes de un data warehouse

2.3.3.1.1. Datos operacionales

Los datos que se obtiene de las bases operacionales usadas para el procesamiento diario de información, necesarias para las diferentes transacciones que usen las aplicaciones de la institución (Ydirín, 2004).

2.3.3.1.2. Datos históricos

Son datos que no son usados regularmente en los sistemas operacionales, debido a que la información que se necesita para el trabajo constante debe ser actualizada reflejando siempre el estado de actividad y el momento actual de la organización, el *data warehouse* permite manejar esta información con los distintos valores que toma en el tiempo suministrando un gran valor para la institución al descubrir tendencias mediante minería de datos (Abella, Cópola, & Olava, 2000).

2.3.3.1.3. Gestor de carga

El gestor de carga realiza todas las operaciones de extracción y carga de datos, se recolecta directamente de las bases operacionales al *data warehouse*, en el cual se puede incluir transformaciones sencillas (Connolly & Begg, 2005).

2.3.3.1.4. Administrador del data warehouse

El administrador es el encargado de todas las operaciones y procesos de manejo de datos en el *data warehouse* como son: (Connolly & Begg, 2005).

- Análisis de datos para determinar su consistencia.
- Generar operaciones de agregación.
- Creación de vistas e índices.
- Copias de seguridad.

2.3.3.1.5. Gestor de consultas

El gestor realiza las operaciones necesarias que están relacionadas con las consultas de los usuarios, direccionando de una manera apropiadas las consultas a sus respectivas tablas (Valencia Arcos & Guevara Lenis, 2007).

2.3.3.1.6. Datos detallados

Todos los datos detallados en el *data warehouse* se almacenan en un esquema de base generalmente en el esquema de estrella, en el cual son agregados y puestos a disposición al siguiente nivel de detalle (Connolly & Begg, 2005).

2.3.3.1.7. Metadatos

Datos acerca de datos, es decir un catálogo en el cual se describe como la información se encuentra estructurada en el *data warehouse* (Martínez, 2008).

La estructura de los metadatos difiere entre cada proceso.

- En los procesos de extracción y carga se utilizan metadatos como un mapa de los datos fuentes, para la ubicación de los datos almacenados.
- Para el proceso de consultas se utiliza los metadatos para dirigir al origen de datos más apropiado (Connolly & Begg, 2005).

2.3.3.1.8. Herramientas de minería de datos

Herramientas que utilizan algoritmos de inteligencia artificial, *machine learning* para descubrir nuevas tendencias o patrones que los usuarios nunca hubiesen sospechado que existen en la institución facilitando la construcción de modelos predictivos (Valencia Arcos & Guevara Lenis, 2007).

2.4. MINERÍA DE DATOS

2.4.1. Definición de minería de datos

La minería de datos es el proceso de aplicar herramientas tecnológicas que permiten extraer información de almacenes de datos, facilitando a las instituciones en el descubrimiento de patrones y nuevas tendencias que los usuarios nunca sospecharon (Monsalve, Aponte, & Hoyos, 2013). Se realiza mediante el uso de modelos o algoritmos de inteligencia artificial, que pueden ser descriptivos o predictivos permitiendo predecir nuevos comportamientos a partir de la información almacenada en el *data warehouse* (Pignani, 2011).

2.4.2. Técnicas de minería de datos

2.4.2.1. Modelo predictivo o de aprendizaje supervisado

El modelo permite predecir características importantes de algún hecho llamado aprendizaje supervisado. Para formar este modelo se utiliza un almacén con datos históricos que posibilite la formación de una muestra de datos llamados grupos de aprendizaje con la capacidad de adaptarse a las pruebas con nuevos modelos de datos (Connolly & Begg, 2005).

2.4.2.1.1. Modelo de clasificación

Es una técnica utilizada para clasificar una clase o atributo predeterminado llamados como variables dependientes, basados en otros atributos llamados variables independientes. Las principales técnicas son: arboles de decisión, redes neuronales, redes de creencias bayesianas (Santamaría, 2010).

2.4.2.1.1.1. Arboles de decisión

Un árbol de decisión mantiene una estructura jerárquica que facilita en la toma de decisiones finales ya que sigue las condiciones que se cumplan desde la raíz del árbol hasta sus hojas (Santamaría, 2010). Las ramas representan salidas, los nodos hojas reflejan a las clases, y el nodo de la parte superior se conoce como raíz (Pazmiño & Suasnavas, 2012).

2.4.2.1.1.2. Redes neuronales

Las redes neuronales es un modelo que aprende de una forma supervisada o no supervisada, en el cual se intenta predecir resultados de ejemplos conocidos comparando con sus predicciones y aprendiendo de sus errores, semejante a una red neuronal biológica (Santamaría, 2010).

2.4.2.1.1.3. Redes bayesianas

La clasificación bayesiana usa la distribución de probabilidades para cuantificar incertidumbre de los datos que van a ser analizados, se utiliza como una buena metodología como inferencia y predicción, una de las desventajas de este método es que no puede realizar un correcto modelo de predicción con pocos datos (Figueroa, 2006).

2.4.2.2. Modelo descriptivo o de aprendizaje no supervisado

Los principales objetivos de este tipo de minería es poder localizar patrones y tendencias descubriendo información para obtener o llevar a cabo alguna acción en beneficio de la institución (Moreno, Quintales, Garcia, & Polo, 2008).

2.4.2.2.1.1. Segmentación

La segmentación se basa en dividir una base de datos en un número desconocidos de segmentos similares, compartiendo las mismas propiedades descubriendo subpoblaciones homogéneas (Connolly & Begg, 2005).

2.4.2.2.1.2. Detección de desviación

El objetivo de esta técnica es identificar los valores atípicos que expresan desviación desde algún conocimiento previo, en esta técnica se utiliza métodos estadísticos como la regresión lineal que facilita la identificación de valores atípicos (Connolly & Begg, 2005).

2.4.3. Minería de datos educacionales

La minería de datos educacionales aplica el uso de algoritmos de inteligencia artificial para detectar patrones de datos educativos, de tal manera los actores principales son estudiantes, docentes y autoridades de una organización académica para identificar actividades, recursos que podrán mejorar el rendimiento académico (Monsalve et al., 2013).

Estas técnicas son usadas para: (Ballesteros Román, Sánchez-Guzmán, & García Salcedo, 2013).

- La comprensión del comportamiento de los estudiantes.
- Analizar el nivel participativo de los estudiantes (aprendizaje – enseñanza).
- Índices de graduación.

2.4.3.1. Principales enfoques

Existen varias técnicas de minería de datos analizadas en la sección 2.4.2 que permiten obtener un conocimiento para la toma de decisiones institucionales.

La Tabla 2 muestra los diferentes enfoques que tiene la minería de datos educacionales.

Tabla 2. Enfoques de la minería de datos educacionales

(García, 2010)

Categoría del método	Objetivo del método	Aplicaciones clave
Predicción	Desarrollo de un modelo que pueda inferir una variable a partir de la combinación de los datos disponibles.	Detección del comportamiento del estudiante. Desarrollo de modelos de dominio, predicción y entendimiento de los resultados académicos de un estudiante.
Agrupamiento	Encontrar conjuntos de datos que se agrupen naturalmente separado el conjunto completo en una serie de categorías.	Descubrimiento de nuevos patrones de comportamiento de estudiantes.
Descubrimiento mediante modelos	Modelado de un fenómeno mediante predicción, agrupamiento o ingeniería del conocimiento, es usado como componente en una futura predicción.	Descubrimiento de relaciones entre comportamiento de estudiantes y sus características o variables contextuales.

2.5. RALPH KIMBALL

2.5.1. Arquitectura

Ralph Kimball define a un *data warehouse* como la reunión de todos los *data marts* de una empresa o institución, siendo una copia de los datos transaccionales de una manera estructurada específicamente para consultas y análisis (Rivadera, 2010). En el cual se tienen las tablas de dimensiones para análisis con sus respectivos atributos, y las tablas de hechos que describen los sucesos del negocio que se quieren analizar.

La metodología de *Kimball* utiliza los diferentes *data marts* conectados entre sí mediante la estructura de bus que contiene todos los elementos conformadas por sus respectivas dimensiones. Esta característica especial

permite construir diferentes *data marts* en cada departamento de acuerdo a sus necesidades.

La Figura 8 describe la arquitectura de *Kimball* en un *data warehouse*.

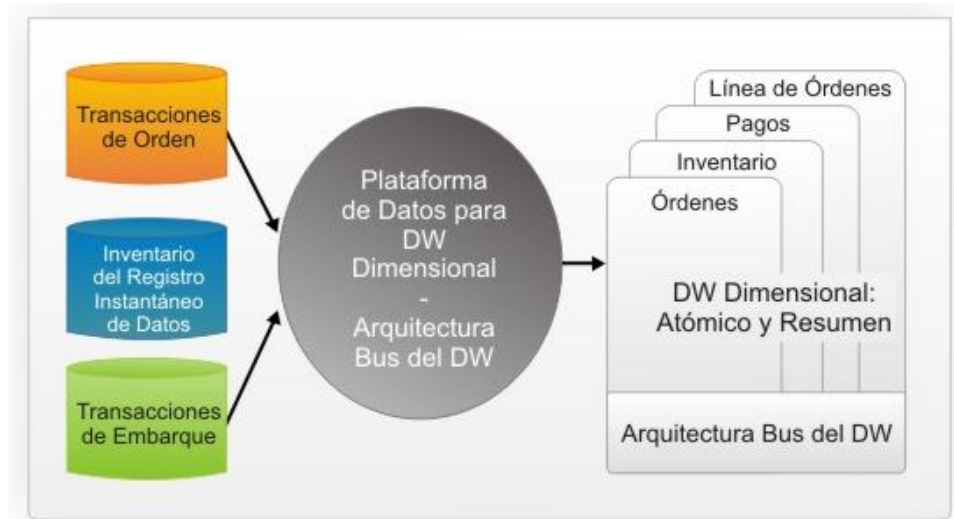


Figura 8. Data Warehouse

(Espinosa, 2010)

2.5.2. Ciclo de vida

Kimball se basa en el ciclo de vida dimensional del negocio el cual se fundamenta en cuatro fases principales (Ilbay, 2009).

- Concentrarse en el negocio.
- Construir una infraestructura de información adecuada.
- Realizar entregas en incrementos significativos.
- Ofrecer una solución completa.

La Figura 9 muestra las diferentes tareas del ciclo de vida de *Kimball* para la construcción de un *data warehouse* desde la planificación hasta su implementación.

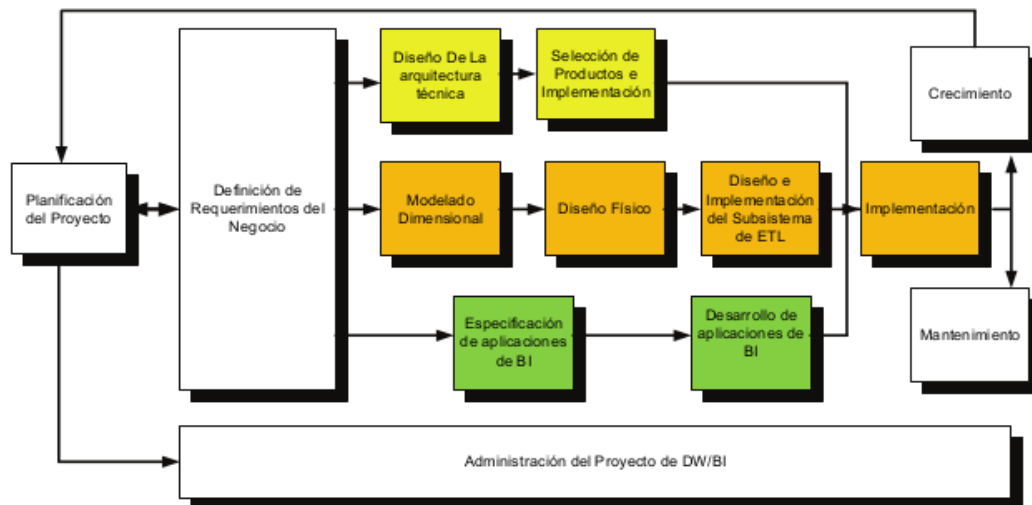


Figura 9. Ciclo de vida *Kimball*

(Rivadera, 2010)

2.5.2.1. Planificación del proyecto

Es el proceso en el cual se determina el propósito del *data warehouse*, sus objetivos específicos, el alcance que tendrá y su factibilidad. En la planificación identificamos las tareas, las asignamos analizando el uso de los recursos del proyecto y su tiempo de duración. La planificación depende de la definición de requerimientos del negocio, debido a que los requerimientos determinan el alcance y recursos del proyecto (Ilbay, 2009).

2.5.2.2. Definición de requerimiento del negocio

La definición de requerimientos se basa en el proceso de interpretar de una manera correcta los diferentes niveles de requerimientos de los usuarios mediante entrevistas, encuestas para poder entender los factores claves del negocio y poder traducirlo en un modelo apropiado.

Según *Kimball* los requerimientos están posicionados en el centro del universo del *data warehouse*, los requisitos deben determinar el alcance del *data warehouse* (qué datos debe contener, cómo debe estar organizado, cada cuánto debe actualizarse, quienes y desde donde accederán, etc.) (Ilbay, 2009).

2.5.2.3. Modelado dimensional

La definición de requerimientos ayuda a identificar los datos necesarios utilizados para el análisis de los usuarios. El modelo dimensional comienza con una matriz, en la cual se determina la dimensionalidad de cada indicador y luego se especifican los diferentes grados de detalle (atributos), dentro de cada concepto del negocio (dimensión), como así también la granularidad de cada indicador (variable o métrica), las diferentes jerarquías que dan forma al modelo dimensional del negocio (Ilbay, 2009).

La Figura 10 muestra el flujo del proceso de dimensión con un proceso iterativo que consta de cuatro pasos iterativos.

1. Elegir el proceso de negocio: Esta fase depende de los requerimientos del negocio descritos en la etapa anterior.
2. Establecer el nivel de granularidad: En este paso se describe el nivel de detalle que tendrá la información en el *data warehouse* de acuerdo a los requerimientos anteriores.
3. Elegir las dimensiones: Las dimensiones se basan en el nivel de detalle que tendrá la información y representa los factores de análisis del negocio para brindar una perspectiva de los atributos de estudio que se asociaran con las tablas de hechos.
4. Identificar medidas y las tablas de hechos: En esta fase se examina las medidas de análisis del negocio que regularmente son atributos expresados en cantidades, tiempo, dinero, sobre las cuales se desea realizar operaciones como promedio, suma, conteo, mínimos, máximos, en función de una o más dimensiones. Las tablas de hechos evidencian los objetos o eventos de análisis del negocio, por lo general los hechos a almacenarse son siempre numéricos y están en relación con las dimensiones.

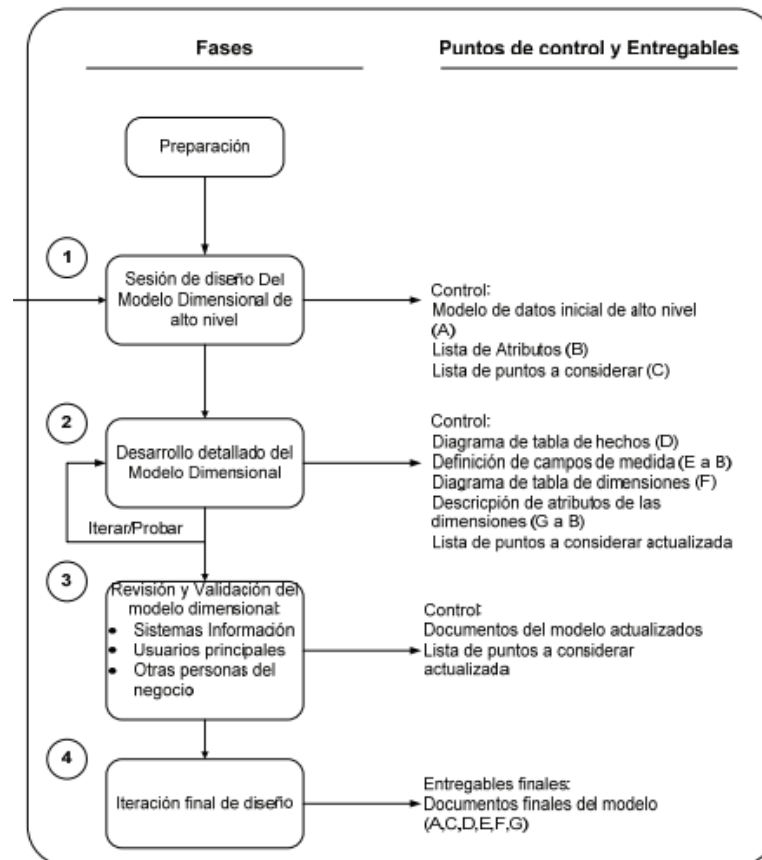


Figura 10. Diagrama del flujo de proceso del proceso dimensional *Kimball*
(Rivadera, 2010)

2.5.2.4. Diseño físico

El diseño físico se basa en la selección de las estructuras necesarias para soportar el diseño lógico de acuerdo con los diferentes requerimientos de la institución, aplicando los diferentes estándares que poseerá el *data warehouse* en construcción.

2.5.2.5. Diseño e implementación del subsistema de Extracción, Transformación, Carga (ETL)

El proceso de extracción, transformación, carga que sus siglas en ingles son ETL comienza con la extracción, en el cual se puede obtener los datos que permitan efectuar la carga en el modelo físico. La transformación es el proceso por el cual los datos se reajustan de acuerdo a las necesidades del negocio

para efectuar una correcta carga en el modelo físico. La carga consta de procesos requeridos para poder poblar el *data warehouse* (Rivadera, 2010).

2.5.2.6. Implementación

La implementación es la unión de toda la tecnología de los procesos anteriores para hacerlo accesibles a los usuarios finales. El correcto funcionamiento depende de la capacitación, la comunicación y el soporte técnico.

2.5.2.7. Mantenimiento y crecimiento de un data warehouse

El *data warehouse* depende de los usuarios del negocio para su crecimiento por esta razón es importante mantener una comunicación constante, con una retroalimentación para poder manejar los nuevos requerimientos de los usuarios y mantener un aumento evolutivo del *data warehouse* (Duque, 2010).

2.5.2.8. Especificaciones de aplicaciones de BI

Una parte fundamental de todo proyecto está en proporcionarles a una gran comunidad de usuarios una forma más estructurada y por lo tanto, más fácil, de acceder al almacén de datos. Proporcionándoles este acceso estructurado a través de lo se llama aplicaciones de inteligencia de negocios (Rivadera, 2010).

Las aplicaciones de BI entregan de una mejor manera los informes y herramientas de análisis como los *dashboard* que son útiles para apreciar de una manera gráfica y analítica la información.

2.5.2.9. Diseño de la arquitectura técnica

El diseño de la arquitectura técnica abarca a todos los procesos y herramientas que se aplican a los datos teniendo una planificación antes de comenzar con la construcción, donde *Kimball* recomienda comenzar por una visión global y dividir las hasta un grado donde realmente se pueda implementarlo.

“La arquitectura técnica se divide en dos partes, el *back room* (la parte interna del *warehouse*) y el *front room* (la cara pública del *warehouse*), interactuando constantemente. Mientras los requerimientos del negocio dicen qué necesitamos hacer, la arquitectura técnica responde el interrogante de ¿Cómo haremos?” (Ilbay, 2009).

2.5.2.10. Selección de productos e implementación

Este proceso se basa en el uso del diseño de la arquitectura técnica para poder seleccionar el *software* que funcionara como el motor de base de datos usado para el *data warehouse* mediante una evaluación de los productos existentes. La Figura 11 muestra cómo realizar la evaluación de los diferentes productos de *software*.

Característica	Peso	Producto A	Producto B	Producto C	...
Capacidades Basicas de ETL					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	85				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	70				
Job Control & Scheduling					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	90				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	85				
Metada & Estandares					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx					
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx					
Info de Vendedor	70				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	85				
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx					
Puntaje Total					
Ranking					

Figura 11. Matriz de evaluación de productos

(Ilbay, 2009)

3. METODOLOGÍA

3. METODOLOGÍA

Para el desarrollo del *data warehouse* se utilizó como guía la metodología de *Ralph Kimball*. Esta metodología permite obtener resultados óptimos con varios *data marts* desde la planificación del proyecto hasta la construcción e implementación

3.1. Planificación y administración del proyecto

Para el desarrollo del proyecto se identificó los diferentes procesos que lleva a cabo el departamento de planificación de la UTE como son:

- Acreditación institucional.
- Planificación institucional.

3.1.1. Definición de requisitos del proyecto

El alcance estuvo en función de la definición de requisitos mediante la realización de varias entrevistas y encuestas a los administrativos del departamento de planificación, para determinar los requerimientos necesarios que fueron utilizados en la construcción del *data warehouse*.

Para la preparación de las entrevistas se analizó las fichas técnicas del CEAACES, la cual es utilizada para el proceso de acreditación institucional y consta de todas las variables e indicadores con sus respectivos criterios para la formulación de cálculos.

El sistema de planificación ayudó para comprender la estructura de las metas y objetivos que se utiliza en la institución, para un proceso de mejoramiento continuo de 5 años, con una visión de preparación conjunta de toda la organización en mira a la acreditación institucional.

Para la creación de un modelo de gestión del conocimiento en el área de planificación se comparó las entrevistas y encuestas, para la determinación de los procesos previos que se llevan a cabo por parte de la unidad, para

conseguir las variables mencionadas anteriormente obteniendo un conocimiento que permita la toma de decisiones institucional.

3.1.2. Modelo dimensional

Para realizar esta fase se analizó el modelo entidad relación que tiene la universidad y debido a la gran cantidad de tablas con falta de documentación, se optó por realizar un modelo de base en *Microsoft Excel*, que por medio de la unidad de nuevas tecnologías de la UTE, con la apoyó de los responsables de cada departamento en la verificación previa, puedan ser actualizados y sirvan como orígenes de datos para poder realizar el proceso de carga, tomando en cuenta las variables e indicadores que ayudan a la planificación y acreditación de la institución.

Se identificó las dimensiones de acuerdo a los requerimientos del proyecto, con la ayuda de los archivos de base, se modeló para que se pueda obtener un mayor dinamismo a la hora de analizar la información obteniendo un diseño que permita ser fácil de utilizar para el usuario.

En las tablas de hechos se identificó las medidas o los valores de cálculo que se utilizan en cada variable que solicita el CEAACES, de acuerdo a requerimientos del proyecto.

Para el diseño se utilizó el modelo de estrella en el *data mart* de planificación encargado de las metas y objetivos institucionales, el *data mart* de acreditación se diseñó con el modelo de constelación de hechos, permitiendo tener una mayor facilidad al momento de generar cubos multidimensionales.

3.1.3. Modelo físico

Se realizó entrevistas a los administradores de base de datos de la universidad, para determinar que licencias son usadas y están disponibles para la creación de la base de datos.

Se identificó que toda la base de datos tiene licenciamiento con *Microsoft*, permitiendo utilizar herramientas de *Business Intelligence*, facilitando

posteriormente la integración de manera nativa en los sistemas institucionales añadiendo una información dinámica y útil al usuario.

3.1.4. Diseño e implementación del proceso ETL

Para esta etapa se utilizó como diseñador de la arquitectura de la base de datos la herramienta *Power Designer*, para la implementación del proceso ETL se empleó las herramientas de *Microsoft SQL Server Data Tools 2012* con *Business Intelligence*.

Los modelamientos de cada *data mart* se realizaron por separado evitando confusiones, debido a la gran cantidad de dimensiones y tablas de hechos que fueron requeridas por el negocio, generando dos *scripts* del modelo físico en SQL para cada *data mart*.

- Carga de datos históricos

Se realizó la carga de dimensiones por primera vez, tomando como base de origen de pruebas los archivos enviados por el ITIC desde el año 2009 hasta el año 2015. Para la tabla de dimensión “tiempo” se realizó una carga preliminar de varias fechas que comprende desde el año 1900 hasta el 2020.

Para las tablas de hechos se realizó preliminarmente el poblado de las dimensiones, para poder obtener sin problemas las asignaciones de sus respectivas claves primarias que correspondan a sus acciones.

- Carga de datos incremental

Para este proceso se realizó pruebas desde la segunda corrida de datos con los procesos ETL implementados en el proyecto, para la extracción y carga se realiza exclusivamente de nuevos datos o que hayan sufrido un cambio en el origen.

- Minería de datos

Para este proceso se realizó un filtrado de datos de acuerdo al análisis que se desea obtener, para el proyecto se desarrollaron análisis con información de

estudiantes tomando como objetivos fechas de culminación de mallas curriculares y estados de graduación.

Se utilizaron herramientas de filtrado ETL con funciones de *Analysis Services*, para poder generar cubos analíticos multidimensionales con la información necesaria, la cual se utilizó para algoritmos de pruebas.

3.1.5. Diseño de la arquitectura técnica

Para esta fase se realizó los diseños preliminares de la estructura que tendrá la base de datos de acuerdo a los requerimientos del negocio. Es importante mantener una constante comunicación con el cliente final en esta etapa, para el diseño de los reportes que serán generados mediante el uso de cubos multidimensionales. Se investigaron herramientas de reportes como son *Pentaho* o *Microsoft Excel*, para tener dos soluciones y comparar las de mejor facilidad de uso con generación de análisis para el departamento.

3.1.6. Selección del producto

Ayudado con entrevistas realizadas a la unidad de nuevas tecnologías de la UTE se realizó una investigación, análisis de factibilidad y pruebas de funcionamiento de las herramientas recomendadas, para establecer criterios unificados y presentar varias propuestas al departamento de planificación, para poder elegir la herramienta a ser utilizada.

3.1.7. Aplicaciones de BI

En esta fase se evaluó el ambiente técnico del departamento para determinar que tecnología es la apropiada hacia la presentación de información al usuario final. Debido a que el departamento maneja como herramienta principal de análisis *Microsoft Excel*.

4. RESULTADOS

4. RESULTADOS

Tomando como base la metodología de *Ralph Kimball* en el desarrollo del proyecto se muestran los resultados desde la planificación hasta la implementación del mismo.

4.1. PLANIFICACIÓN Y ADMINISTRACIÓN DEL PROYECTO

4.1.1. Definición requisitos del proyecto

El área de planificación mostró un gran interés para obtener conocimiento que les permita fortalecer la gestión universitaria, mediante la toma de decisiones oportunas e indicadores que añadan valor a la información.

Para determinar los requisitos se realizó varias entrevistas y encuestas con los interesados, que determinaron que la información necesaria a administrar es de profesores, estudiantes, variables de acreditación y las diferentes metas institucionales.

La Figura 12 indica la información necesaria en el área de planificación para la creación del *data warehouse*.

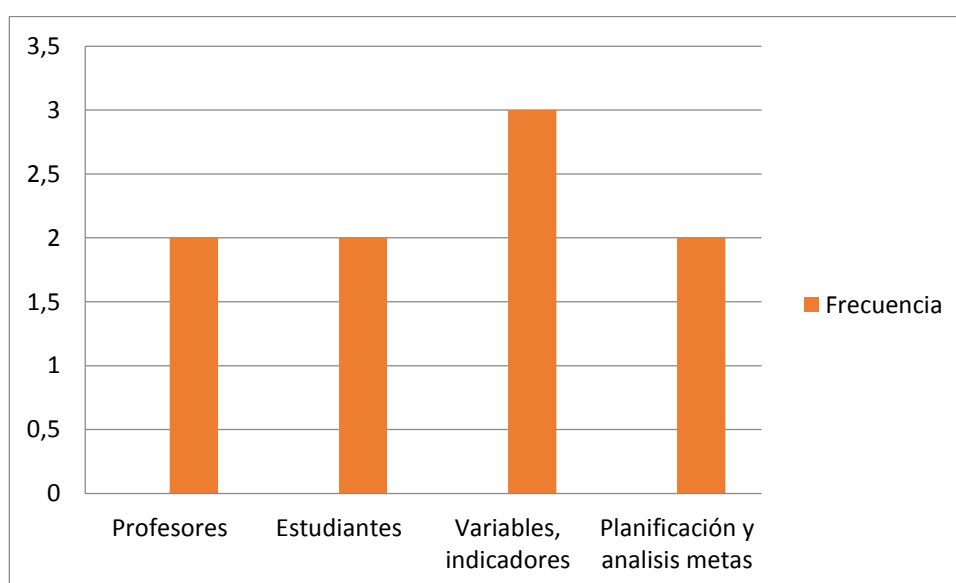


Figura 12. Información necesaria en el área de planificación

4.1.2. Modelo dimensional

La Tabla 3 indica el esquema de la base de origen que alimentara al *data warehouse*.

Tabla 3. Esquema Base de origen

Nombre del esquema	Nombre del esquema en la base de origen
Planificación	Planificación.xls
Docentes	PAC-DEFINITIVO.xls
Estudiantes pregrado	Estudiantes.xls
Estudiantes posgrado	EstudiantesPos.xls
Estudiantes primer nivel	EstudiantesPriNiv.xls
Asignaturas pregrado	AsignaturasDoc.xls
Asignaturas posgrado	AsignaturasDocPos.xls
Infraestructura	Ancho_Banda.xls
Biblioteca	Biblioteca.xls
Campus	CampusUte.xls
Tiempo	DimensionTiempo.xls
Publicaciones	Libro_articulo_docente.xls
Presupuesto	Presupuesto_Universitario.xls
Proyectos de vinculación	PROYECTOS_VINCULACION.xls
Proyectos de investigación	Proyectos_Investigacion.xls

4.1.2.1. Nivel de granularidad

Para esta fase se definió el nivel de granularidad consiguiendo una independencia de datos, y la información pueda tener un nivel de detalle amplio permitiendo soportar ciertas consultas que beneficien a la institución.

4.1.2.2. Identificar las dimensiones y mapeo de datos

La Tabla 4 muestra el mapeo de datos de las dimensiones y tablas de hechos de planificación institucional.

Tabla 4. Mapeo de datos del *data mart* de planificación institucional

Tablas de dimensiones y hechos	Anexo
Tabla de hechos porcentaje cumplimiento meta.	Anexo 1. Tabla de hechos porcentaje meta
Dimensión inversión.	Anexo 2. Dimensión Inversión
Dimensión código kpi.	Anexo 3. Dimensión Código Kpi
Dimensión unidades.	Anexo 4. Dimensión Unidades
Dimensión objetivo unidad.	Anexo 5. Dimensión Objetivo Unidad
Dimensión tiempo.	Anexo 6. Dimensión Tiempo
Dimensión meta mes.	Anexo 7. Dimensión Código Meta Mes
Dimensión meta a cumplirse.	Anexo 8. Dimensión Meta Cumplirse
Dimensión meta indicador logro.	Anexo 9. Dimensión Meta Indicador Logro
Dimensión inversión gasto.	Anexo 10. Dimensión Inversión gasto
Dimensión meta logros alcanzados.	Anexo 11. Dimensión Logros Alcanzados
Dimensión objetivo institucional.	Anexo 12. Dimensión Objetivo Institucional
Dimensión meta objetivo unidad.	Anexo 13. Dimensión Objetivo Unidad
Dimensión meta plan de mejora	Anexo 14. Dimensión Meta Plan de Mejora
Dimensión meta retroalimentación.	Anexo 15. Dimensión Meta Retroalimentación
Dimensión meta usuario unidad.	Anexo 16. Dimensión Meta Usuario Unidad

La Tabla 5 indica el mapeo de datos de las dimensiones y tablas de hechos de acreditación institucional.

Tabla 5. Mapeo de datos del *data mart* de acreditación institucional

Tablas de dimensiones y hechos	Anexo
Dimensión biblioteca.	Anexo 17. Dimensión Biblioteca
Dimensión campus.	Anexo 18. Dimensión Campus
Dimensión aulas.	Anexo 19. Dimensión Aulas
Dimensión presupuesto.	Anexo 20. Dimensión Presupuesto
Dimensión materias.	Anexo 21. Dimensión Materias
Dimensión carreras.	Anexo 22. Dimensión Carreras
Dimensión facultad.	Anexo 23. Dimensión Facultad
Dimensión docente.	Anexo 24. Dimensión Docente
Dimensión tiempo dedicación.	Anexo 25. Dimensión Tiempo dedicación
Dimensión período sabático.	Anexo 26. Dimensión Período Sabático
Dimensión ingreso concurso.	Anexo 27. Dimensión Ingreso Concurso
Dimensión sub área último título cuarto nivel del docente.	Anexo 28. Dimensión Sub área Último Título Cuarto Nivel
Dimensión período académico.	Anexo 29. Dimensión Período
Dimensión cursando estudios.	Anexo 30. Dimensión Cursando Estudios
Dimensión cargo autoridad.	Anexo 31. Dimensión Cargo Autoridad
Dimensión último título cuarto nivel del docente.	Anexo 32. Dimensión Último Título Cuarto Nivel
Dimensión categoría personal académico.	Anexo 33. Dimensión Categoría Personal Académico

Tabla 5. Mapeo de datos del *data mart* de acreditación institucional
continuación

Tablas de dimensiones y hechos	Anexo
Dimensión del último título de tercer nivel de docente.	Anexo 34. Dimensión Último Título Tercer Nivel
Dimensión sexo.	Anexo 35. Dimensión Sexo
Dimensión modalidad del estudiante.	Anexo 36. Dimensión Modalidad
Dimensión colegio.	Anexo 37. Dimensión Colegio
Dimensión discapacidad.	Anexo 38. Dimensión Discapacidad
Dimensión estado civil.	Anexo 39. Dimensión Estado Civil
Dimensión localidad.	Anexo 40. Dimensión Localidad
Dimensión estudiantes primer nivel.	Anexo 41. Dimensión Estudiantes Primer Nivel
Dimensión estudiantes posgrado.	Anexo 42. Dimensión Estudiante Posgrado
Dimensión publicación artículos.	Anexo 43. Dimensión Publicación Artículos
Dimensión libro capítulo.	Anexo 44. Dimensión Libro Capitulo
Dimensión estudiantes pregrado.	Anexo 45. Dimensión Estudiantes Pregrado
Dimensión revista indexada.	Anexo 46. Dimensión Revista Indexada
Dimensión proyectos de vinculación.	Anexo 47. Dimensión Proyectos de Vinculación
Dimensión proyectos de investigación.	Anexo 48. Dimensión Proyectos de Investigación
Tabla de hechos estudiantes primer nivel.	Anexo 49. Tabla de Hechos Estudiantes Primer Nivel
Tabla de hechos internet.	Anexo 50. Tabla de Hechos Internet
Tabla de hechos de los proyectos de vinculación.	Anexo 51. Tabla de Hechos Proyectos de Vinculación

Tabla 5. Mapeo de datos del *data mart* de acreditación institucional
continuación

Tablas de dimensiones y hechos	Anexo
Tabla de hechos del presupuesto universitario.	Anexo 52. Tabla de Hechos Presupuesto Universitario
Tabla de hechos de los estudiantes de posgrado.	Anexo 53. Tabla de Hechos Estudiantes Posgrado
Tabla de hechos biblioteca.	Anexo 54. Tablas de Hechos Biblioteca
Tabla de hechos aulas de la institución.	Anexo 55. Tabla de Hechos Aulas
Tabla de hechos libros artículos.	Anexo 56. Tablas de Hechos Libros Artículos
Tabla de hechos período sabático.	Anexo 57. Tabla de Hechos Periodo Sabático
Tabla de hechos de docentes.	Anexo 58. Tabla de Hechos Docentes
Tabla de hechos de estudiantes pregrado.	Anexo 59. Tabla de Hechos Estudiantes Pregrado
Tabla de hechos de los proyectos de investigación.	Anexo 60. Tabla de Hechos Proyectos de Investigación
Tabla de hechos de las asignaturas de pregrado.	Anexo 61. Tabla de Hechos Asignatura Docentes Pregrado
Tabla de hechos cursando estudios de los docentes.	Anexo 62. Tabla de Hechos Cursando Estudios

4.1.3. Modelo físico

Para esta fase se utilizó como base de datos *Microsoft SQL Server 2008 R2*, que son paquetes de *software* licenciados en la institución, para la generación de tablas con *scripts* del *data warehouse* se utilizó herramientas de modelamiento de *Sybase*.

4.1.4. Diseño e implementación del proceso ETL

4.1.4.1. Extracción

Para llevar a cabo la primera fase del proceso ETL, se selecciona los diferentes orígenes de fuentes que posee la base de datos transaccional. Se selecciona como origen de datos *Excel*, que funcionan como fuentes de datos previamente reunidas por la unidad de nuevas tecnologías de la institución.

La Figura 13 muestra la fuente de datos seleccionada.



Figura 13. Origen de datos

A continuación se debe asignar las columnas de entrada que están en el modelo base de *Excel*, asignar a las columnas de salida del componente de flujo de datos como se muestran en la Figura 14.

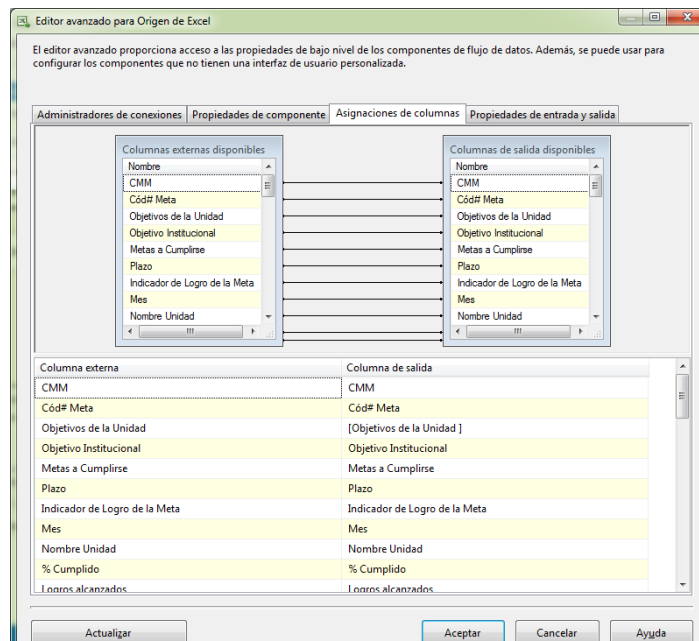


Figura 14. Asignación de columnas de entrada y salida

4.1.4.2. Transformación

Para esta fase se realizó un tratamiento con los valores nulos transformándoles en valores establecidos con los interesados, para una mejor visualización de la información. Para el tratamiento se utilizó las transformaciones mediante columnas derivadas, conversiones de datos, uniones, agrupaciones, ordenamientos, que proporciona la herramienta *SQL Server Data Tools 2012*. La Figura 15 muestra el tratamiento que se dio a los diferentes valores nulos que se encontró en las fuentes de datos mediante la opción de *REPLACENULL*.

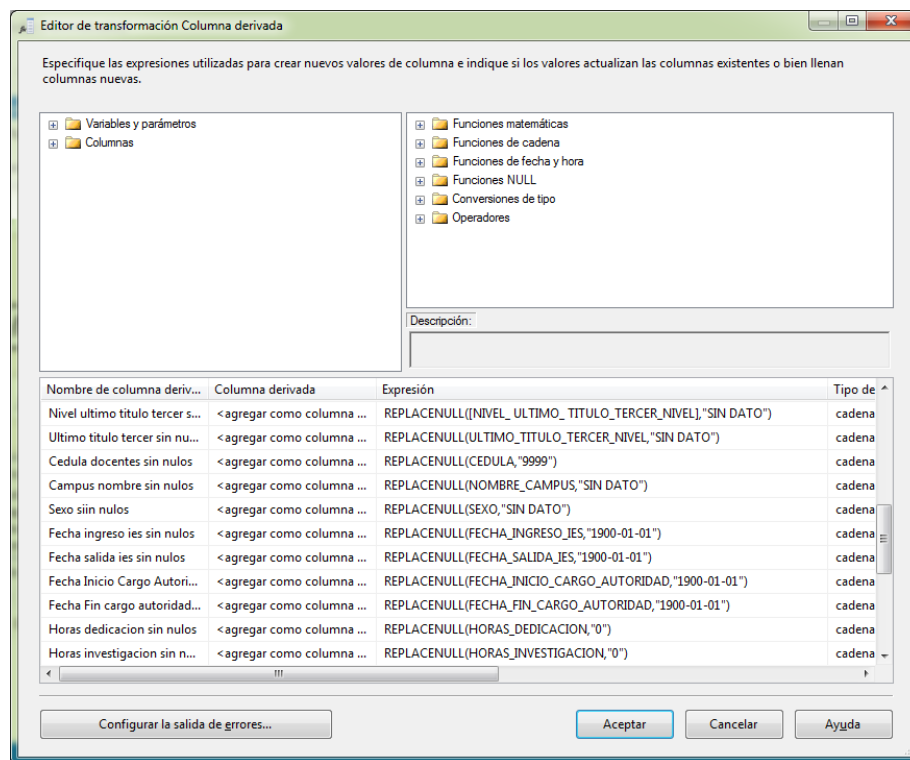


Figura 15. Transformación de valores nulos

En la Figura 16 se observa el proceso que se utilizó en las cadenas extensas con más de 255 caracteres, mediante el uso de la opción *SUBSTRING*. La Figura 17 indica la transformación que se dio a los metadatos para la conversión con su tipo de dato correspondiente.

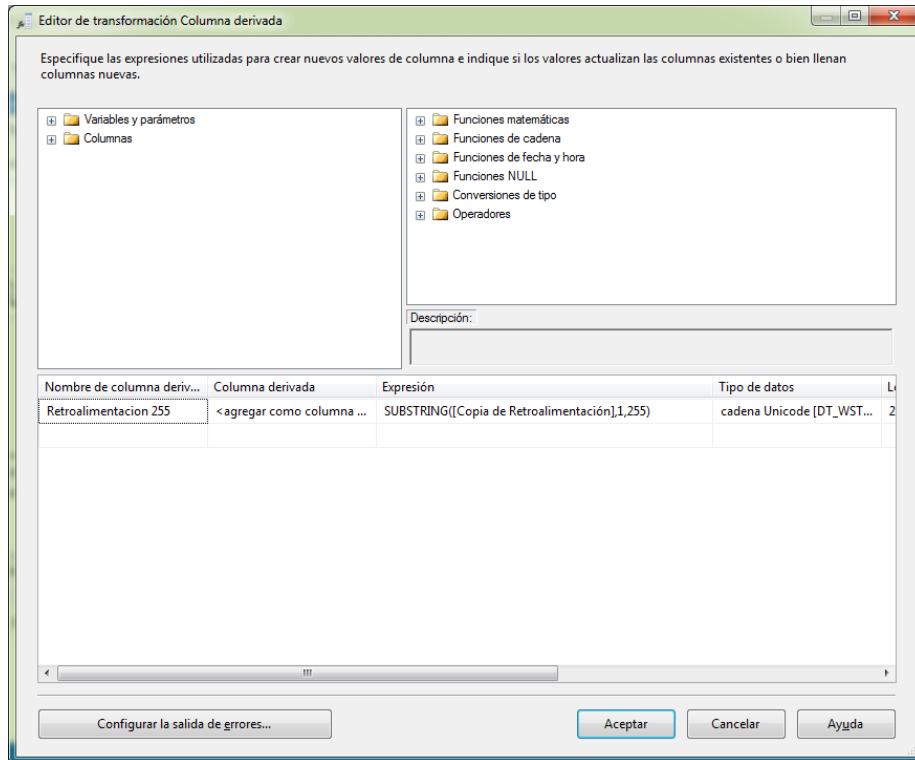


Figura 16. Transformación de cadenas extensas

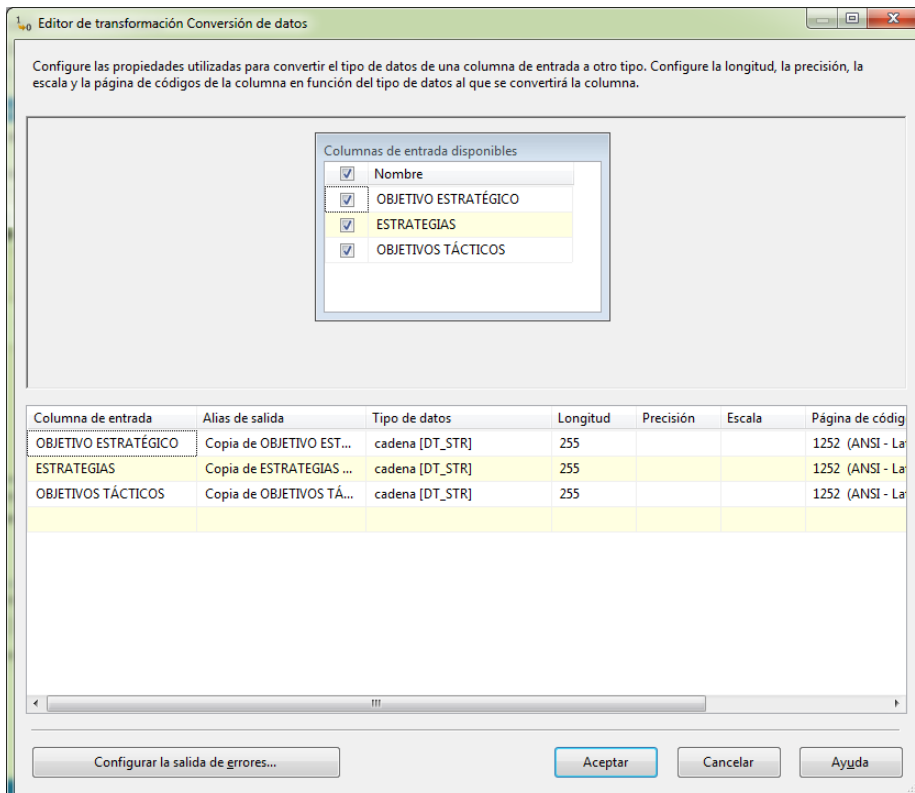


Figura 17. Transformación de conversión de datos

4.1.4.3. Carga

Para este proceso se utilizó como destino las diferentes tablas de dimensiones y hechos del *data warehouse*, asignando las entradas y salidas disponibles con sus correspondientes tipos de datos. La Figura 18 muestra las diferentes columnas de asignaciones disponibles, para poder ser insertadas en el *data warehouse*.

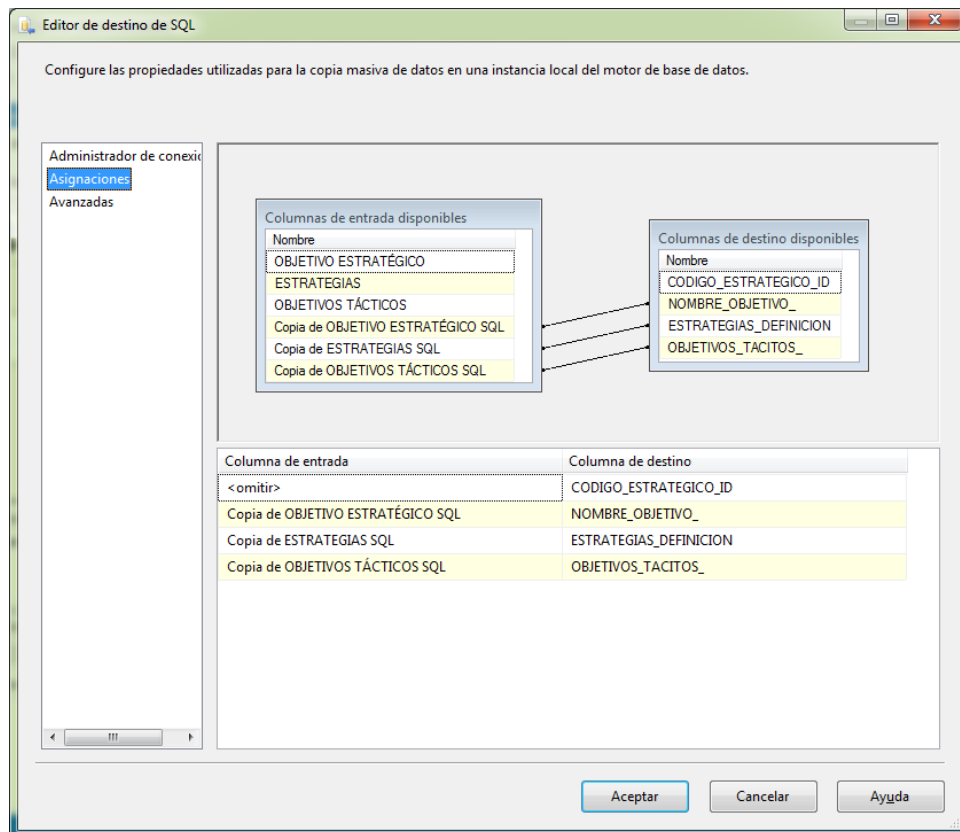


Figura 18. Editor de destino de SQL

4.1.5. Diseño de la arquitectura técnica

4.1.5.1. Back Room

En esta fase se implementa toda la parte del proceso de extracción, transformación, carga (ETL) desde el origen de datos hasta el destino final SQL para poblar el *data warehouse*.

4.1.5.2. Front Room

Para esta etapa se utilizó las herramientas de *Microsoft Excel* como análisis de los resultados obtenidos, debido a que la mayoría de los informes del departamento son realizados en el *software* mencionado anteriormente.

4.1.6. Selección del producto

La Tabla 6 muestra la evaluación realizada con información técnica de cada producto, adecuada a las necesidades de la institución.

Tabla 6. Evaluación de productos

Criterio	Peso General	Pentaho		Microsoft BI		Cognos		Oracle	
		Punto	Peso	Punto	Peso	Punto	Peso	Punto	Peso
Integración con diferentes base de datos	0,05	5	0,25	4	0,2	4	0,2	4	0,2
Generación de reportes BI	0,2	4	0,8	4	0,8	4	0,8	5	1
Integración con Microsoft Office	0,05	3	0,15	5	0,25	3	0,15	4	0,2
Análisis de datos	0,2	4	0,8	5	1	5	1	5	1
Asistencia técnica	0,05	1	0,05	5	0,25	4	0,2	4	0,2
Control de acceso y seguridad	0,05	4	0,2	5	0,25	5	0,25	5	0,25
Carga de datos y rendimiento del proceso ETL	0,1	4	0,4	5	0,5	5	0,5	5	0,5
Integración con diferentes sistemas operativos	0,05	5	0,25	3	0,15	3	0,15	3	0,15
Interfaz de uso amigable	0,05	4	0,2	5	0,25	2	0,1	3	0,15
Total			3,1		3,65		3,35		3,65

Para el desarrollo del proyecto se seleccionó las herramientas de *Microsoft* debido a que la institución tiene como base de datos transaccional *Microsoft SQL Server 2008 R2*, contando con las diferentes licencias para ser utilizadas en todo el proyecto y debido a su muy buena puntuación en los diferentes procesos realizados en la evaluación de productos.

4.1.7. Aplicaciones de BI

Las aplicaciones que serán utilizadas por los usuarios finales son complementos añadidos al programa *Microsoft Excel*, que permite realizar minería de datos utilizando el *data warehouse* implementado como se muestra en la Figura 19.



Figura 19. Complemento de minería de datos en *Microsoft Excel*

4.1.8. Implementación

En esta etapa se utilizó los manuales técnicos con los requerimientos de *hardware* de cada producto.

Requerimientos de *hardware* de *Microsoft SQL Server 2008 R2*:

- Procesador *Intel Xeon* de 2,0 GHz o más.
- 8 GB o máximo del sistema operativo.
- La capacidad de disco duro para un *data warehouse* en GB se muestra en la Tabla 7.

Tabla 7. Tamaño de disco duro

	OLTP (Base operacional)	Data warehouse
Tamaño BD en GB	1	$(OLTP * 2 - OLTP) * 100$

Requerimientos de *hardware* para *SQL Server Data Tools 2012*:

- Procesador a 1,6 GHz o superior.
- 1 GB de RAM.
- 10 GB (NTFS) de espacio disponible en disco duro.

Requerimientos de *hardware* para *Pentaho*:

- Procesador de 4 Cores.
- Memoria Ram 8 GB. (4 Gb para *Pentaho Data Integration* Dedicado, 4 *Servidor Apache Tomcat*).
- Espacio en disco duro 80 GB.

En la Tabla 8 se observa la cantidad de datos en filas que fueron procesadas y colocadas en la base de datos en construcción.

Tabla 8. Total de filas procesadas en los *data marts*

Archivo	Cantidad de filas de datos
Ancho_Banda	3
Aulas_Ute	268
Biblioteca	4
CampusUte	4
Tiempo	29587
libro_articulo_docente	5
Presupuesto_Universitario	25
Proyectos_Investigacion	1
Proyectos_Vinculacion	128
Salinas AsignaturasDoc	1572
Sto AsignaturasDoc	4859
Uio AsignaturasDoc	34807
PAC-2014 DEFINITIVO UIO	983
PAC-2014 DEFINITIVO STO	171
PAC-2014 DEFINITIVO SAL	47
PAC-2015 DEFINITIVO UIO	998
PAC-2015 DEFINITIVO STO	232
Salinas EstudiantesPregrado	3998
Sto Estudiantes	22069
Uio Estudiantes	146808
Sto EstudiantesPos	143
Uio EstudiantesPos	1904
Sal EstudiantesPriNiv	635
Sto EstudiantesPriNiv	2832
Uio EstudiantesPriNiv	18118
Planificacion	2862
TOTAL EN ARCHIVOS	273063
TOTAL DE DATOS EN LOS DATA MARTS	
Total fact tables acreditación	234092
Total dimensiones acreditación	129322
Total dimensiones planificación	12765
Total fact tables planificación	2862
TOTAL DE FILAS	379041

4.1.9. Mantenimiento y crecimiento

Para el proceso de mantenimiento se realiza un diagnóstico del resultado de ejecución con cada flujo de datos, para que las tablas de hechos puedan ser almacenadas de una manera correcta y sin problemas. En esta etapa para un correcto funcionamiento el responsable de recopilar los archivos fuentes en

Microsoft Excel, debe verificar que estén acordes a las plantillas establecidas anteriormente y eliminar filas o columnas que se encuentren en blanco.

Para la fase de crecimiento se entregarán los archivos fuentes de cada flujo de datos con sus correspondientes diagramas físicos y códigos *scripts* utilizados en el proyecto.

El proyecto desarrollado con *SQL Server Data Tools 2012* se encuentra dividido por paquetes, uno por dimensiones y otro por tablas de hechos, los cuales están totalmente comentados en cada flujo de datos con su respectivo proceso de ETL, para cada tabla del *data warehouse* que servirá como guía para un crecimiento futuro.

4.2. MODELO DE GESTIÓN DE CONOCIMIENTO DE LA UNIVERSIDAD

Para realizar el modelo de gestión de conocimiento en el área de planificación se ejecutó un análisis de los procesos que se llevan a cabo en el departamento para obtener los diferentes indicadores y su utilización en la planificación y la acreditación institucional mediante entrevistas a varios departamentos, en el cual manejan información pertinente para planificación y determinar el proceso que llevan a cabo para extraer y enviar sus respectivas variables.

Para el desarrollo de la propuesta del modelo de gestión se pudo contar con la ayuda de un académico de intercambio de la Universidad de Berlín, se formó un equipo de trabajo que apoyó con las entrevistas y sugerencias de modelos utilizados en su institución.

4.2.1. Departamento de Biblioteca

Se procedió a entrevistar al director del departamento de biblioteca, en el cual se maneja 3 variables cuantitativas en el proceso de acreditación. En la entrevista se determinó que la extracción de información de la primera variable lo ejecuta mediante un proceso manual, ayudado por hojas electrónicas de

Excel manteniendo evidencias fotográficas de los espacios en cada biblioteca con *Microsoft Word*.

La segunda y tercera variable se extrae del Sistema de Automatización de Biblioteca UTE (SABUTE) mediante la ayuda de técnicos de la respectiva unidad. La Tabla 9 muestra cómo se extrae la información.

Tabla 9. Variables del departamento de biblioteca

	Variables	Extracción de información
1	Espacios Biblioteca	Conteo manual para tener evidencias fotográficas en <i>Microsoft Word</i> y registro en <i>Microsoft Excel</i> del total.
2	Libros impresos	SABUTE.
3	Número de ingresos bibliotecas	SABUTE.

4.2.2. Instituto de transferencia de tecnologías

Se realizó la entrevista al director del departamento de Investigación y Transferencia de Tecnologías, en el cual se maneja 4 variables cuantitativas para el proceso de acreditación que son enviadas al área de planificación. En la entrevista se pudo determinar que no existe ningún mecanismo automatizado para poder obtener la información de las variables, y que se envía una hoja electrónica a los encargados de las diferentes publicaciones para que puedan ser llenadas con sus respectivas evidencias. La Tabla 10 muestra cómo se extrae información de la unidad.

Tabla 10. Variables del departamento de Investigación y Transferencia de Tecnologías

	Variables	Extracción de información
1	Publicaciones Latindex	Se lleva registro mediante <i>Microsoft Excel</i> .
2	Publicaciones seminarios	Se lleva registro mediante <i>Microsoft Excel</i> .
3	Publicaciones indexadas	Se lleva registro mediante <i>Microsoft Excel</i> .
4	Libros	Se lleva registro mediante <i>Microsoft Excel</i> .

4.2.3. Instituto de informática y computación

Se realizó la entrevista al director del ITIC, debido a que las variables de estudiantes y docentes representan el 70 % de la acreditación institucional. En la entrevista se determinó que la información de estudiantes es almacenada en el sistema de información SICAF contando con más de 5000 tablas, debido a la falta de documentación técnica, la información de estudiantes necesaria para la acreditación es enviada al departamento de planificación mediante un proceso intermedio de consolidación previa de estudiantes. La información de docentes es enviada al departamento de recursos humanos que se encarga de confirmar los datos y formar un *pack* de docentes unificado, el cual es reenviado al departamento de planificación.

4.2.4. Modelo de conocimiento del área de planificación

Después de analizar los procesos que actualmente maneja la institución para generar conocimiento, se determinó que existen ciertas deficiencias en el uso de los sistemas que provee la universidad para mantener una información consolidada en el *data warehouse* en construcción, afectado principalmente por los procesos repetitivos de consolidación de datos que existen en los diferentes departamentos. La Figura 20 muestra el modelo de gestión de conocimiento generado para la unidad.

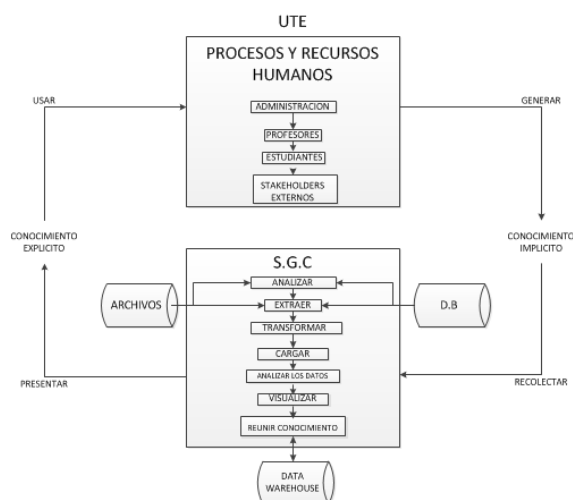


Figura 20. Propuesta de modelo de gestión de conocimiento para planificación UTE

Los principales procesos y recursos humanos en la universidad empiezan por el personal administrativo, profesores, estudiantes y los *stakeholders* externos que son las empresas con las cuales se tiene convenios y quienes generan un conocimiento implícito.

Actualmente este conocimiento es recolectado mediante los sistemas informáticos de la universidad, hojas de cálculo, procesadores de texto que algunas veces son extraídos de los propios sistemas como una ayuda para el análisis de procesos complicados.

Para el sistema de gestión de conocimiento que se desea crear se utilizará la información generada por los sistemas de la universidad extraídos mediante hojas de cálculo, el cual será transformado y almacenado en un *data warehouse*. El *data warehouse* guardará toda la información para generar un conocimiento explícito útil el cual será visualizado por los interesados y apoyará en toma de decisiones y creación de nuevo conocimiento.

4.3. DESARROLLO DEL DATA WAREHOUSE

4.3.1. Desarrollo de los data mart

Para el proyecto se diseñaron dos *data mart*, debido a que el departamento de planificación maneja la siguiente información:

- La planificación de objetivos y metas anuales de todos los departamentos de la universidad con una duración de cinco años y con una revisión cada inicio y fin del periodo académico.
- Indicadores y variables que forman parte de la acreditación institucional que se lleva a cabo cada cinco años, se rige a cinco criterios actualmente aprobados por el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES) que son academia, eficiencia académica, investigación, organización e infraestructura.

4.3.1.1. Características de los data mart

El *data mart* de planificación cuenta con las siguientes características:

- 14 tablas de dimensiones de información del departamento.
- 1 tabla de dimensiones de tiempo.
- 1 tabla de hechos o *fact table*.
- 1 tabla actualización.
- Modelo de diseño tipo estrella.

El *data mart* de acreditación cuenta con las siguientes características:

- 32 tablas de dimensiones de información del departamento.
- 1 tabla de dimensiones de tiempo.
- 14 tablas de hechos o *fact tables*.
- 1 tabla actualización.
- 1 tabla de minería de datos.
- Modelo de diseño constelación de hechos.

4.3.1.2. Implantación de los data mart

Para la implantación de los *data mart* se utilizó herramientas de *business intelligence* de *Microsoft*, en el cual se generó diferentes proyectos para cada *data mart* realizando como proceso preliminar la carga de las dimensiones, para después proceder a colocar sus respectivas claves primarias en las tablas de hechos.

La Figura 21 muestra la ejecución de los flujos de datos en *SQL Server Data Tools 2012*.

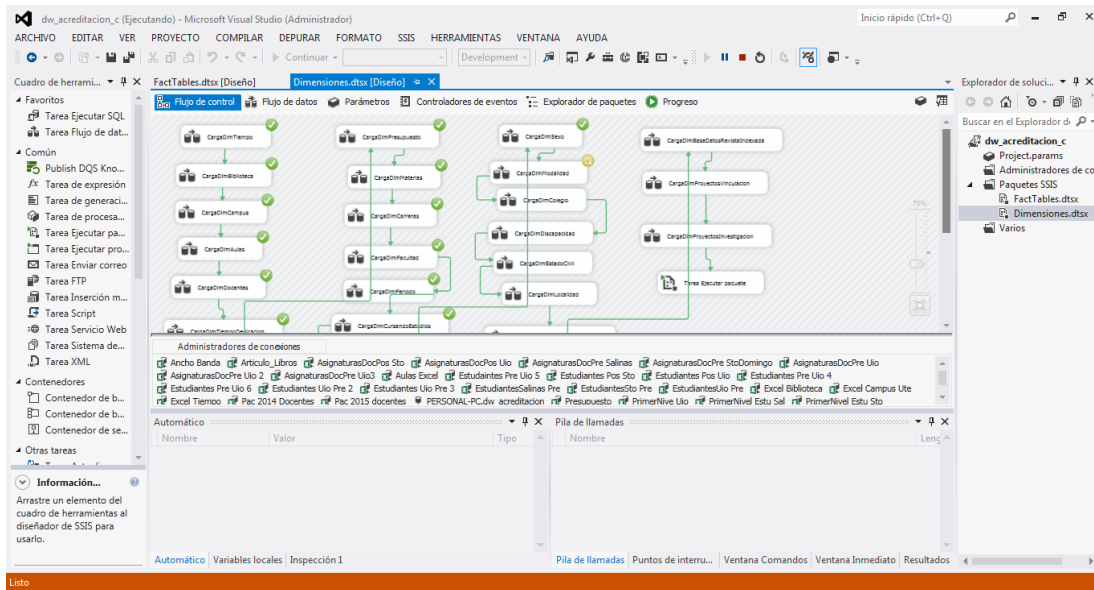


Figura 21. Flujo de datos en dimensiones

En la Figura 22 se observa el desarrollo de flujo de datos en SQL Server Data Tools 2012.

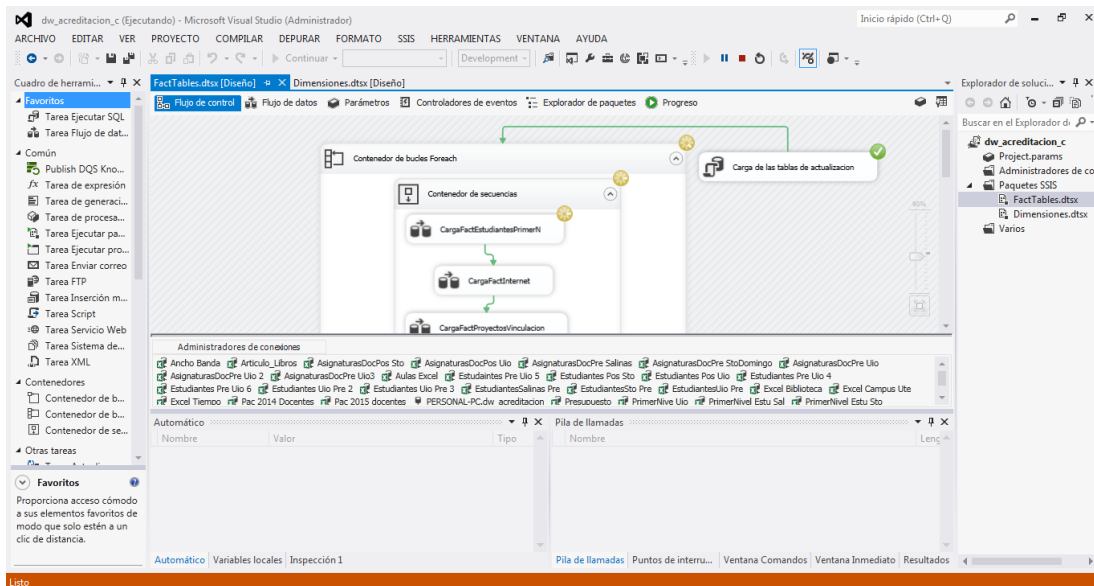


Figura 22. Flujo de datos en tablas de hechos

4.3.1.3. Pruebas de funcionamiento en los data mart

En el transcurso de los experimentos que se realizó a los *data mart*, se localizaron algunas fallas que se tienen directamente en los orígenes de datos de pruebas como los siguientes: Los campos de datos vienen escritos

implícitamente “NULL”, en el cual al momento de comparar si es un dato nulo, pasa esa condición debido a que tiene un dato implícito anteriormente. La Figura 23 muestra que los campos código_carrera vienen escritos “NULL” desde los archivos de base.

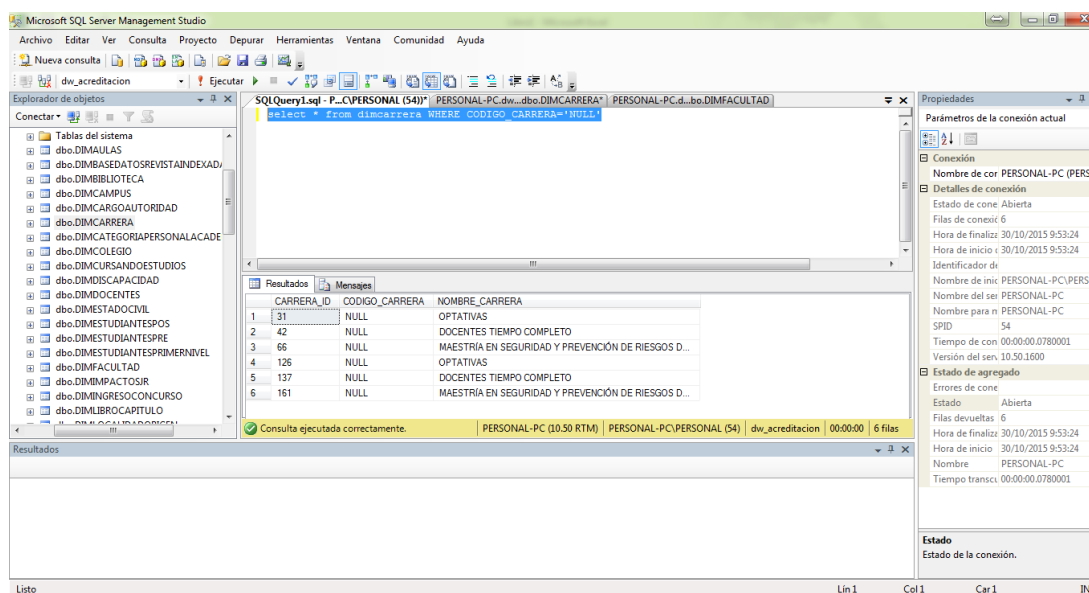


Figura 23. Error de campo “NULL” localizados en los datos de carrera

Se localizó que en los campos de facultades no se tiene unificado en las base de origen en los diferentes campus por ejemplo: En el campus de Santo Domingo se tiene la facultad de “POSTGRADO”, en el cual fue unificado y cambiado por “POSGRADO” para todos los campus.

En la Figura 24 se observa los datos de la facultad del campus Santo Domingo en la base de pruebas.

fecha_inicio	fecha_fin	fecha_egres	fecha_gradu	anio	periodo	facultad	carrera	NOMBRE_CAI
21/03/2009	25/09/2009	25/09/2009	22/02/2010	2009	MAR 2009 - S	POSTGRADO	DIPLOMADO	STO
21/03/2009	25/09/2009	25/09/2009	22/02/2010	2009	MAR 2009 - S	POSTGRADO	DIPLOMADO	STO
21/03/2009	25/09/2009	25/09/2009	22/02/2010	2009	MAR 2009 - S	POSTGRADO	DIPLOMADO	STO
21/03/2009	25/09/2009	25/09/2009	22/02/2010	2009	MAR 2009 - S	POSTGRADO	DIPLOMADO	STO
21/03/2009	25/09/2009	25/09/2009	22/02/2010	2009	MAR 2009 - S	POSTGRADO	DIPLOMADO	STO
21/03/2009	25/09/2009	25/09/2009	22/02/2010	2009	MAR 2009 - S	POSTGRADO	DIPLOMADO	STO

Figura 24. Error de escritura en las base de datos de origen

Se identificó en los archivos de origen de pruebas que existen fechas de nacimiento con valores de campos nulos, en donde se cambió por la fecha “1900-01-01” como se observa en la Figura 25.

NOMBRES	SEXO	NACIONALIDAD	FECHA_NACIMIENTO
GIOVANNY N	HOMBRE	ECUATORIAN	NULL
CHRISTIAN N	HOMBRE	POR CONFIR	NULL

Figura 25. Fechas de nacimiento de docentes con valores nulos

4.3.1.3.1. Generación de cubos OLAP

En *SQL Server Data Tools 2012* se creó un nuevo proyecto para la generación de cubos como se muestra en la Figura 26.

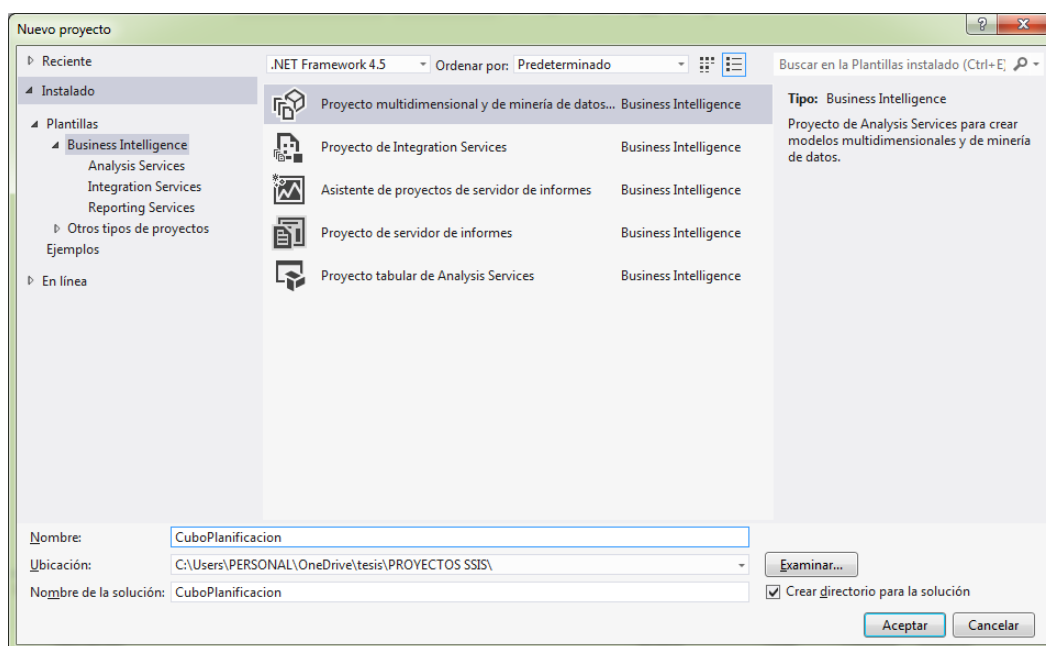


Figura 26. Pantalla de creación de un proyecto multidimensional

Se crea un nuevo origen de datos como muestra la Figura 27, se selecciona la conexión de la base de datos a la cual se desea conectar como se observa en la Figura 28. Se crea una vista a partir de las tablas relacionales como indica la Figura 29.

La Figura 30 muestra el asistente usado para la selección de las tablas a usarse.

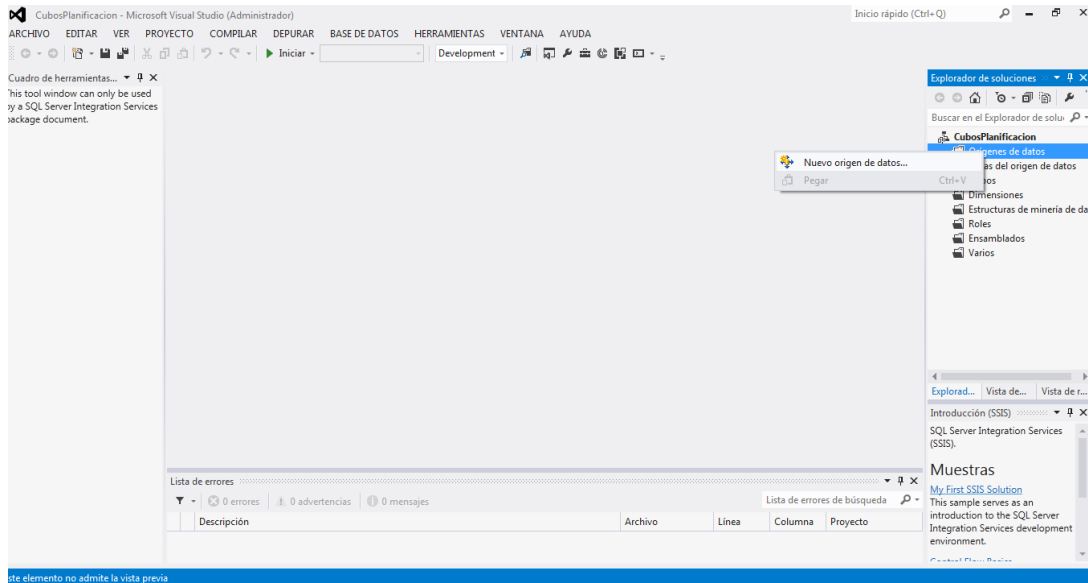


Figura 27. Pantalla de selección de nuevos orígenes de datos

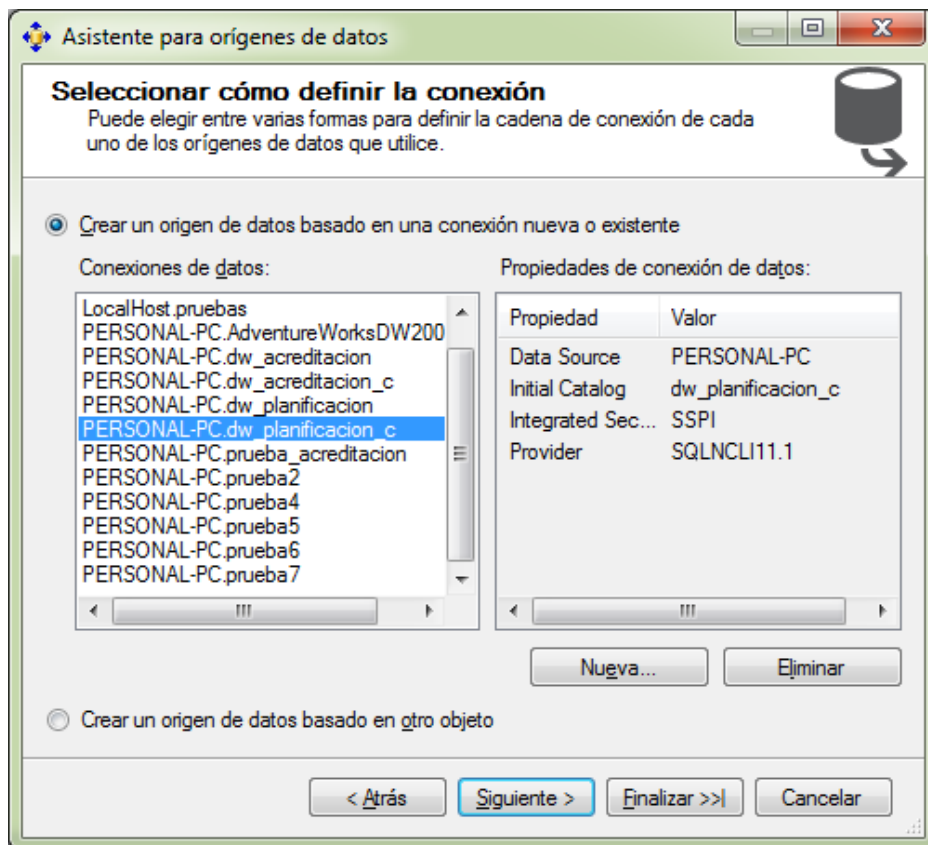


Figura 28. Pantalla de una nueva conexión

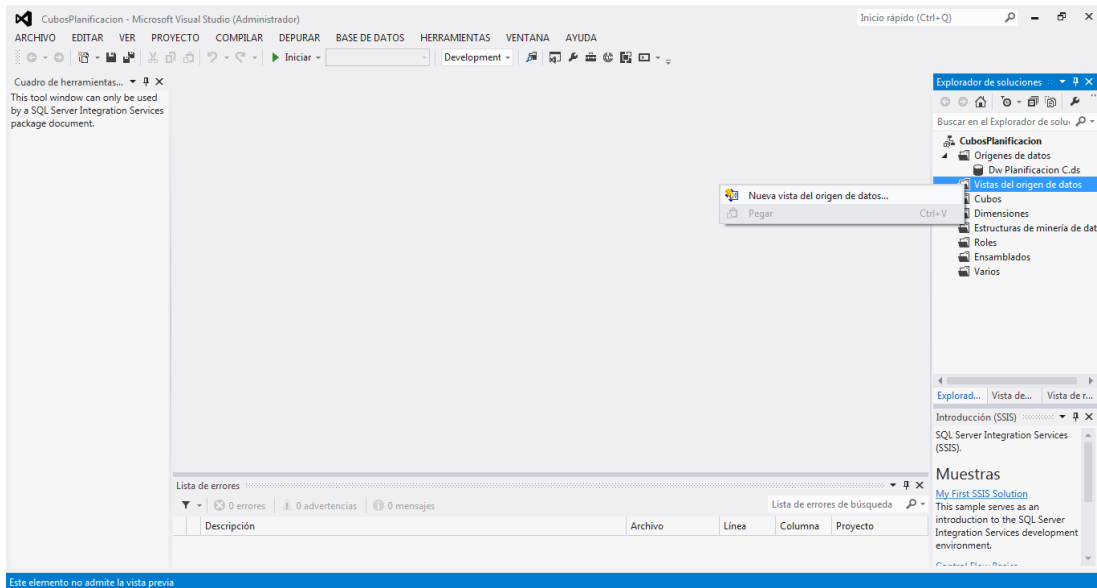


Figura 29. Pantalla de nuevas vistas de origen de datos

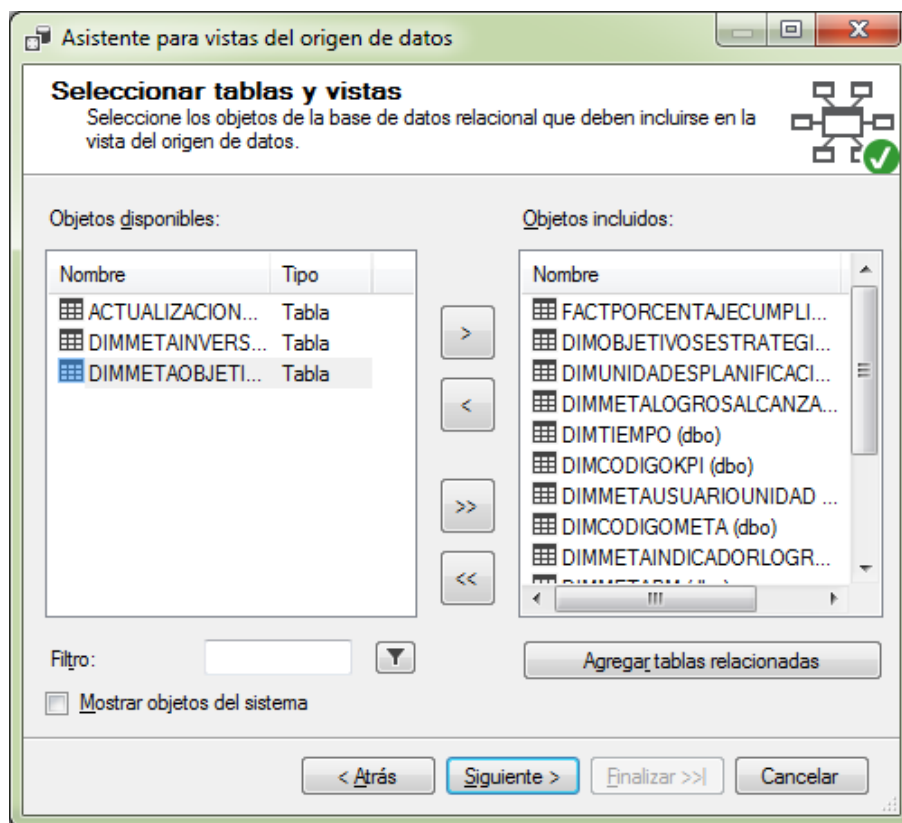


Figura 30. Pantalla de selección de tablas para crear la vista

La Figura 31 muestra la generación de un nuevo cubo usando el asistente, en la Figura 32 se selecciona la tabla en la cual se encuentra el grupo de medidas, y las dimensiones a utilizar como se observa en la Figura 33.

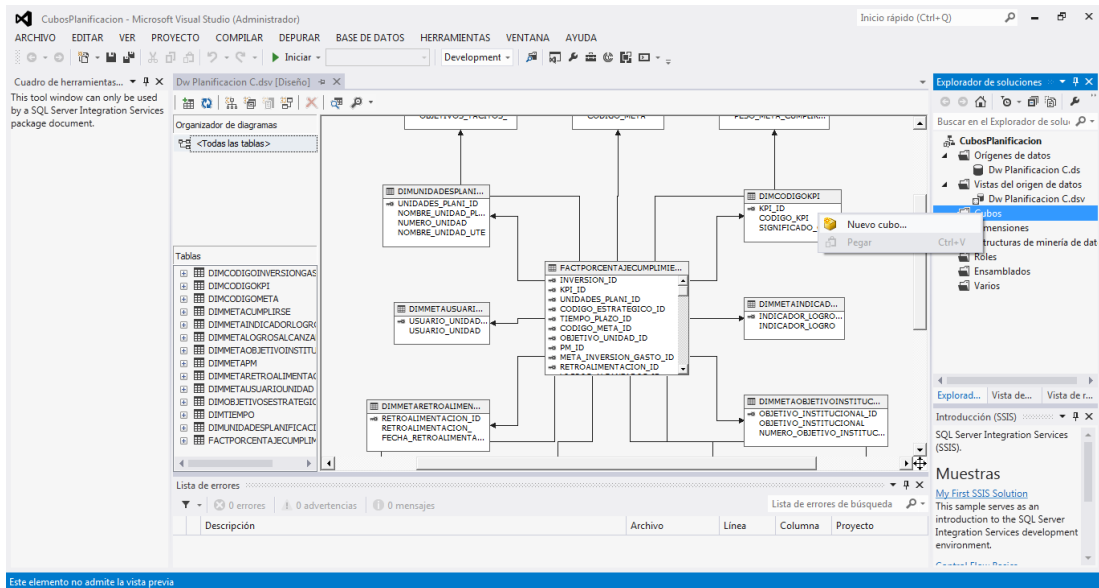


Figura 31. Pantalla de un nuevo cubo

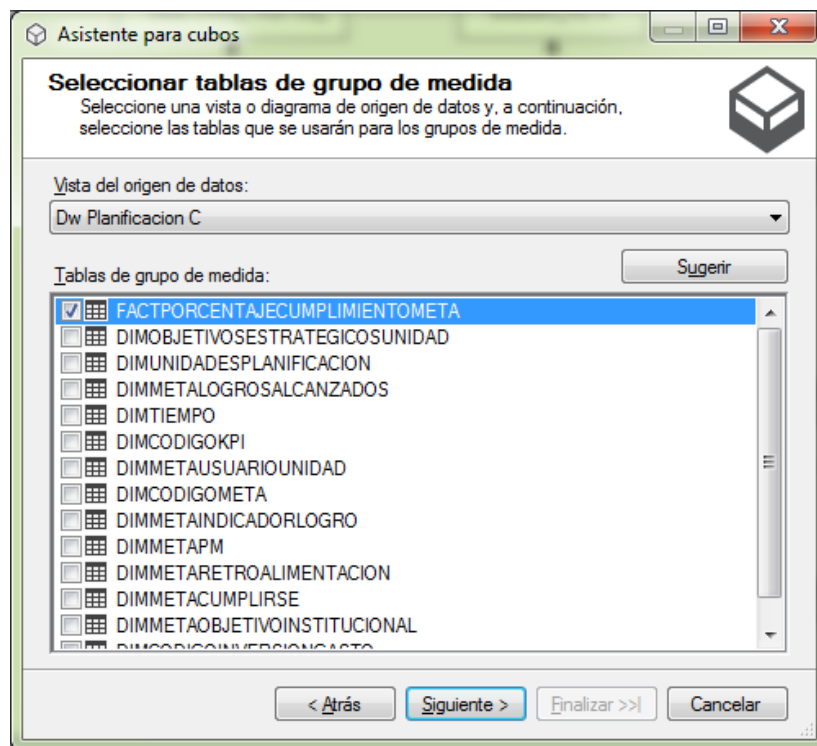


Figura 32. Pantalla del asistente para cubos grupo de medidas

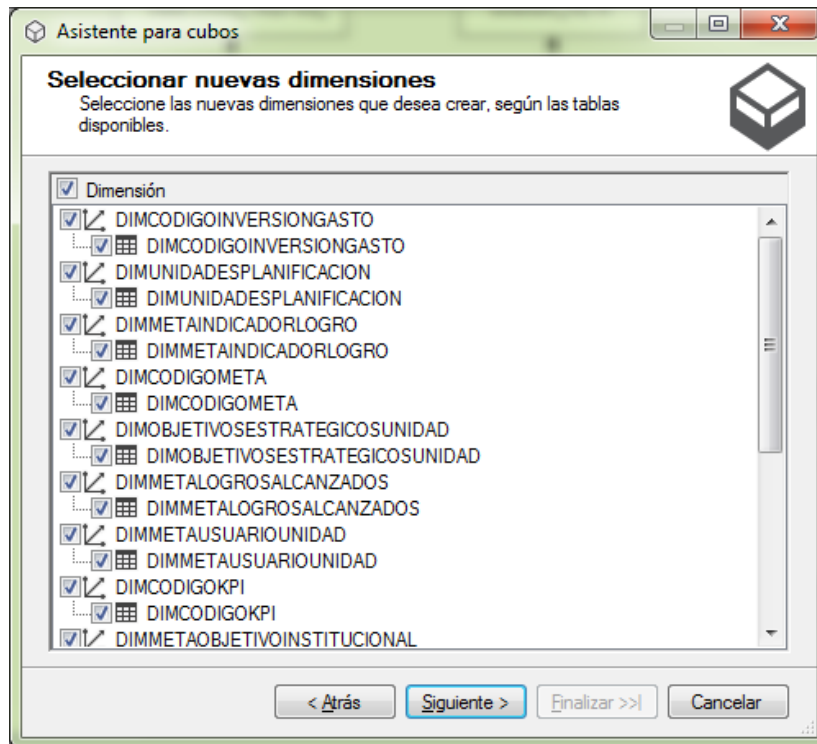


Figura 33. Pantalla asistente para cubos dimensiones

En la Figura 34 se observa la pantalla que permite procesar el cubo para su posterior uso.

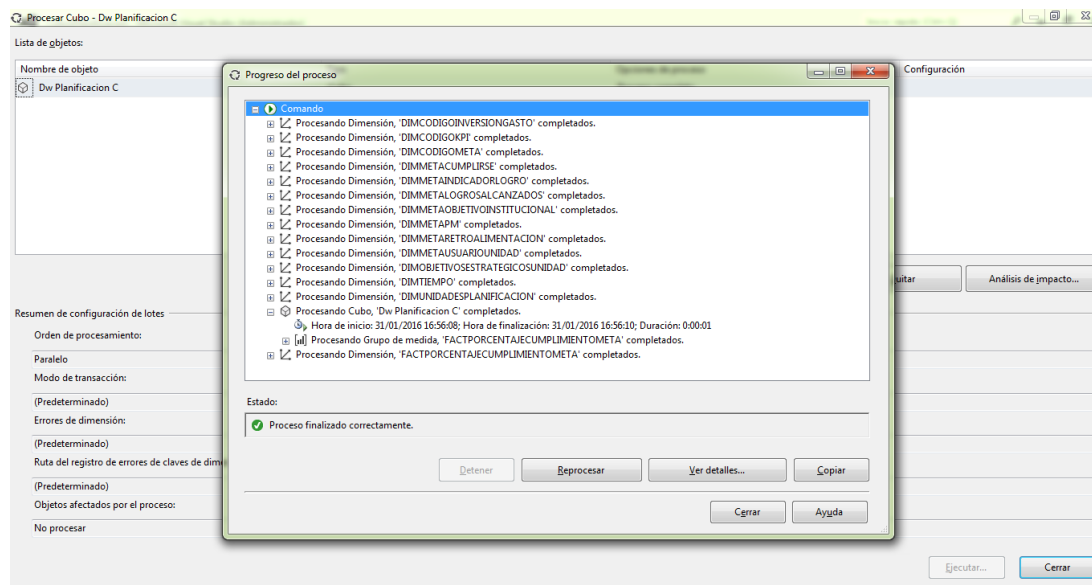


Figura 34. Pantalla de procesamiento del cubo

4.4. EVALUACIÓN Y DESPLIEGE

Para la evaluación del funcionamiento de los *data mart* se utiliza los cubos creados mediante la herramienta de *Microsoft Excel*, permitiendo crear reportes muy rápidos para el usuario final. En la Figura 35 podemos observar la conexión que se realiza al cubo creado anteriormente.

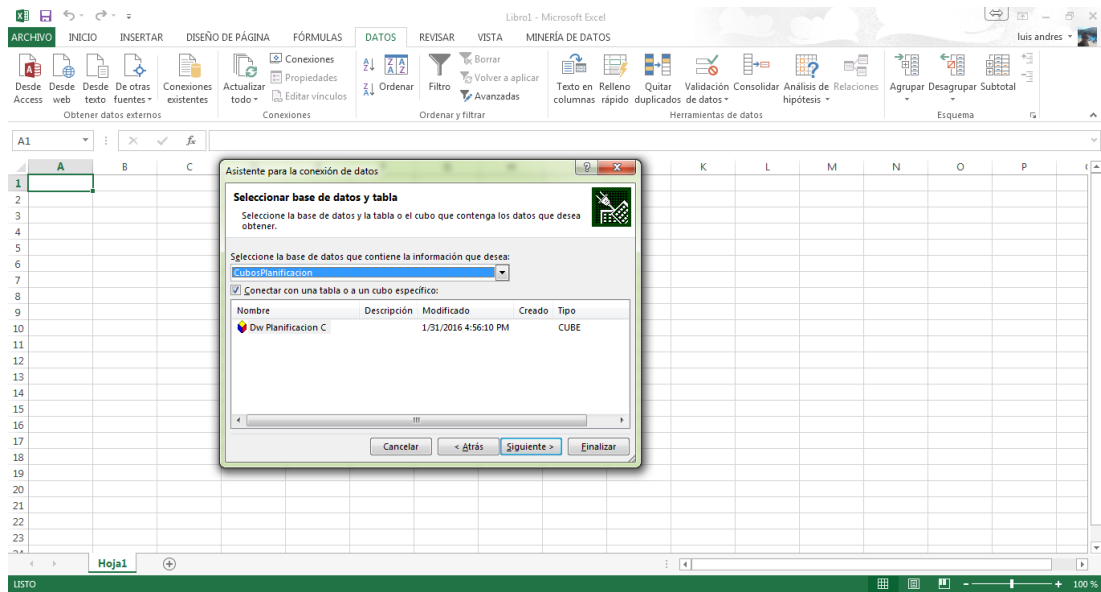


Figura 35. Pantalla de Microsoft Excel de conexión al cubo

Se realizó el diseño de informes de acuerdo a lo que se desea obtener. En la Figura 36, indica el número de metas correspondientes a cada año de acuerdo a su código KPI, con la respectiva unidad que lo maneja. Las metas son estructuradas de acuerdo a criterios que el departamento estipule convenientes para el mejoramiento continuo en el proceso de acreditación.

En la Figura 37 se observa el informe del año 2013 y 2014 de las metas pertenecientes por códigos KPI que tiene la institución con una reducción mínima de metas para el año 2014.

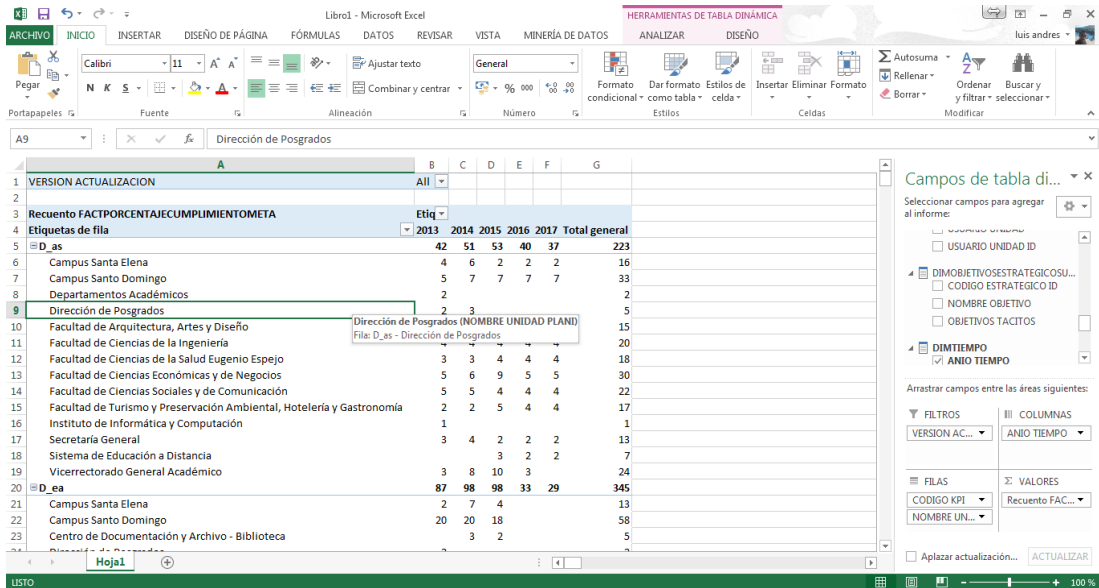


Figura 36. Pantalla informes de planificación por códigos KPI y por nombres de unidad

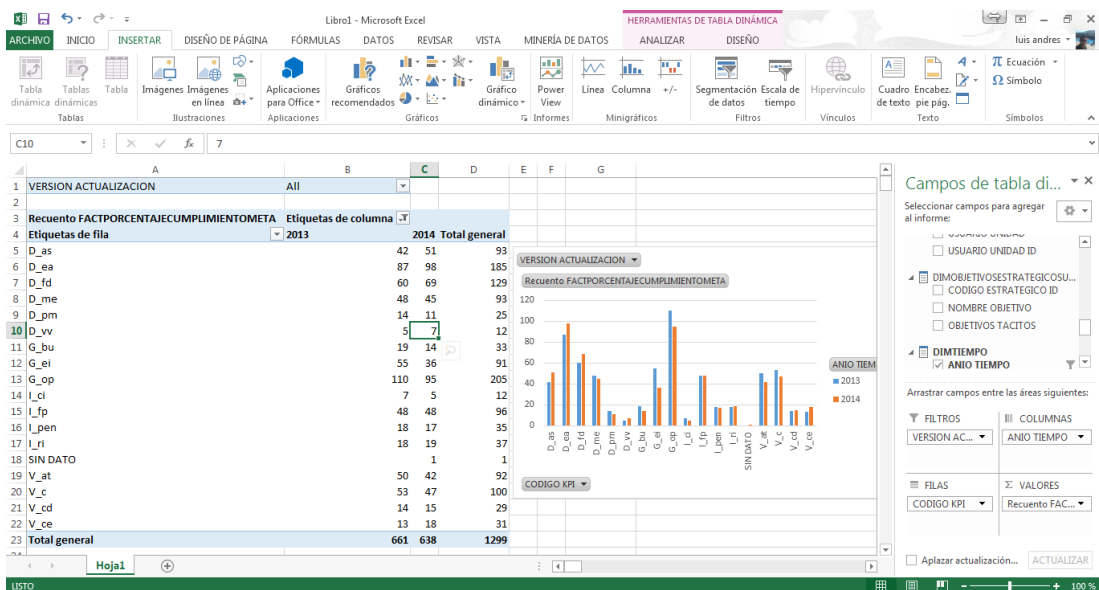


Figura 37. Pantalla informe de planificación por códigos KPI con gráficos

En la Figura 38 muestra que existen metas que no llegaron a cumplirse en su totalidad, a pesar de que en el año 2014 se redujo la cantidad de metas comparado con el año anterior.

La Figura 39 indica la cantidad de graduados desde el año 2010 hasta el año 2015, se aprecia que el sexo femenino tiene un mayor número de graduados

en la facultad de educación a distancia. El sexo masculino tiene un mayor número de graduados en la facultad de ciencias de la ingeniería.

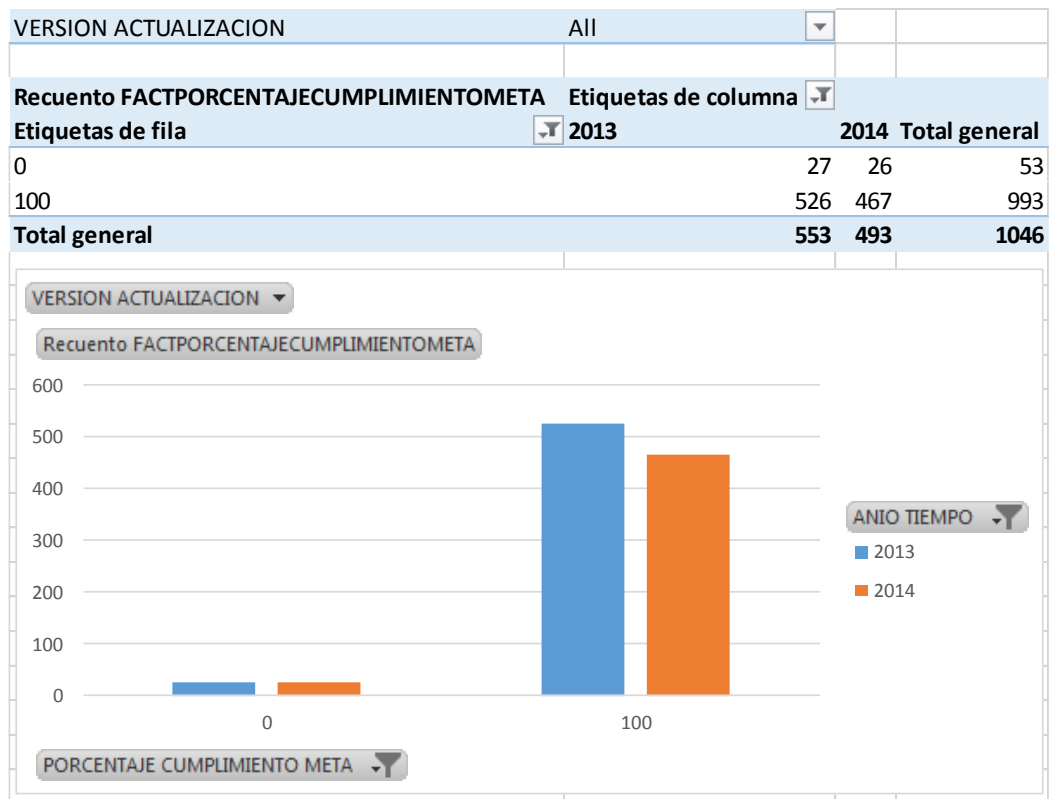


Figura 38. Evaluación de metas del año 2013 y 2014 con un cumplimiento de 0 y 100%

Recuento FACTESTUDIANTESPRE		Etiquetas de columna						
Etiquetas de fila		2010	2011	2012	2013	2014	2015	Total general
FEMENINO		110	960	2372	4387	4573	7110	19512
ARQUITECTURA, ARTES Y DISEÑO		4	56	110	213	195	372	950
CIENCIAS AGROPECUARIAS Y DESARROLLO RURAL				3	43	17	46	109
CIENCIAS DE LA INGENIERÍA		7	66	203	412	597	1043	2328
CIENCIAS DE LA SALUD		27	72	196	140	281	124	840
CIENCIAS ECONÓMICAS Y NEGOCIOS		13	205	421	1263	1299	2231	5432
CIENCIAS SOCIALES Y COMUNICACIÓN		28	83	232	490	326	645	1804
EDUCACIÓN A DISTANCIA		16	329	936	1196	1156	1807	5440
TURISMO, HOTELERÍA Y GASTRONOMÍA		15	149	271	630	702	842	2609
MASCULINO		81	566	1344	2458	2884	5068	12401
ARQUITECTURA, ARTES Y DISEÑO			14	59	105	117	152	447
CIENCIAS AGROPECUARIAS Y DESARROLLO RURAL				3	21	128	49	381
CIENCIAS DE LA INGENIERÍA		11	88	326	614	1033	1908	3980
CIENCIAS DE LA SALUD		14	49	114	124	150	55	506
CIENCIAS ECONÓMICAS Y NEGOCIOS		10	108	226	589	677	1476	3086
CIENCIAS SOCIALES Y COMUNICACIÓN		25	71	110	120	144	257	727
EDUCACIÓN A DISTANCIA		12	136	308	370	355	485	1666
TURISMO, HOTELERÍA Y GASTRONOMÍA		9	97	180	408	359	555	1608
Total general		191	1526	3716	6845	7457	12178	31913

Figura 39. Estudiantes graduados por año, por sexo, por facultad

En la Figura 40 se observa un gráfico del *data mart* de acreditación y se puede analizar la cantidad de estudiantes matriculados, por años y por facultad. Se observa que la facultad de ciencias de la ingeniería, económicas y negocios, educación a distancia son las de mayor número de estudiantes en la institución.

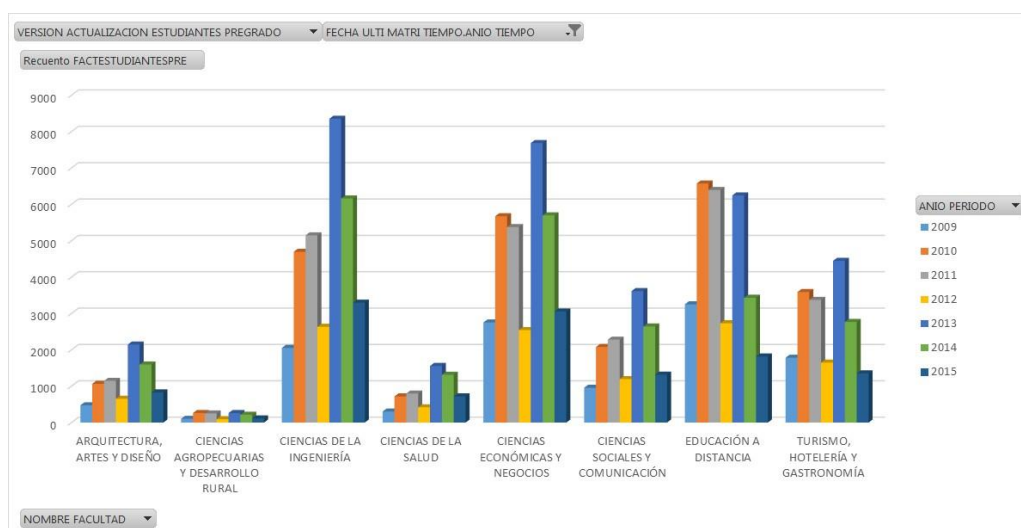


Figura 40. Estudiantes matriculados por año

En la Figura 41 se observa un gráfico de la cantidad de estudiantes por modalidad de estudios, se percibe que la institución tiene una mayoría de estudiantes en la modalidad presencial.

En el año 2013 la facultad de ciencias de la ingeniería alcanzó un gran número de estudiantes, para el siguiente año no lo pudo mantener disminuyendo el número de alumnos matriculados.

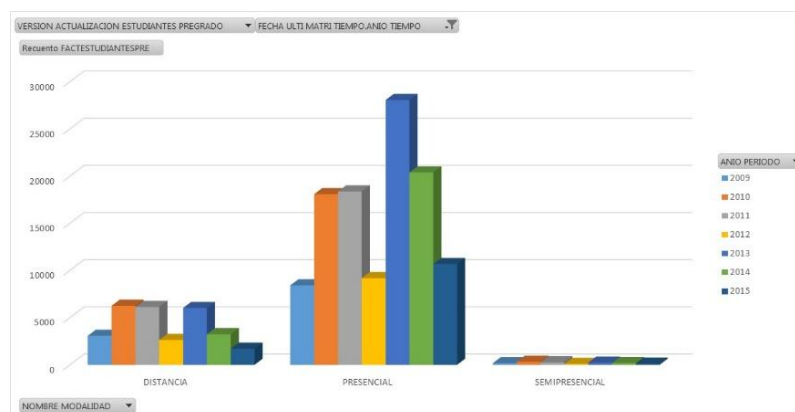


Figura 41. Estudiantes matriculados por modalidad

Para la evaluación de minería de datos se utilizó como herramienta de análisis *Microsoft Excel* con un *Add-In* especial de *Datamining*, que facilita el manejo al usuario final. Se creó una tabla de análisis de estudiantes para guardar los datos que serán utilizados por *Microsoft Excel* para la minería.

Se realizaron dos minerías de datos, la primera analiza a los estudiantes de pregrado que terminaron todas sus asignaturas de estudio y no se han graduado, o ya se han graduado. Para la segunda prueba se utilizó solo a estudiantes de Ingeniería Informática y Ciencias de la Computación, desde el semestre septiembre 2009 hasta julio 2015.

Para utilizar minería de datos en *Microsoft Excel* es necesario crear una instancia o un cubo en *SQL Analysis Services*, para habilitar las funciones analíticas de minería de datos integradas. La Figura 42 muestra la configuración de una nueva instancia en *Analysis Services*, para permitir crear modelos de datos temporales que serán usados en *Microsoft Excel*.

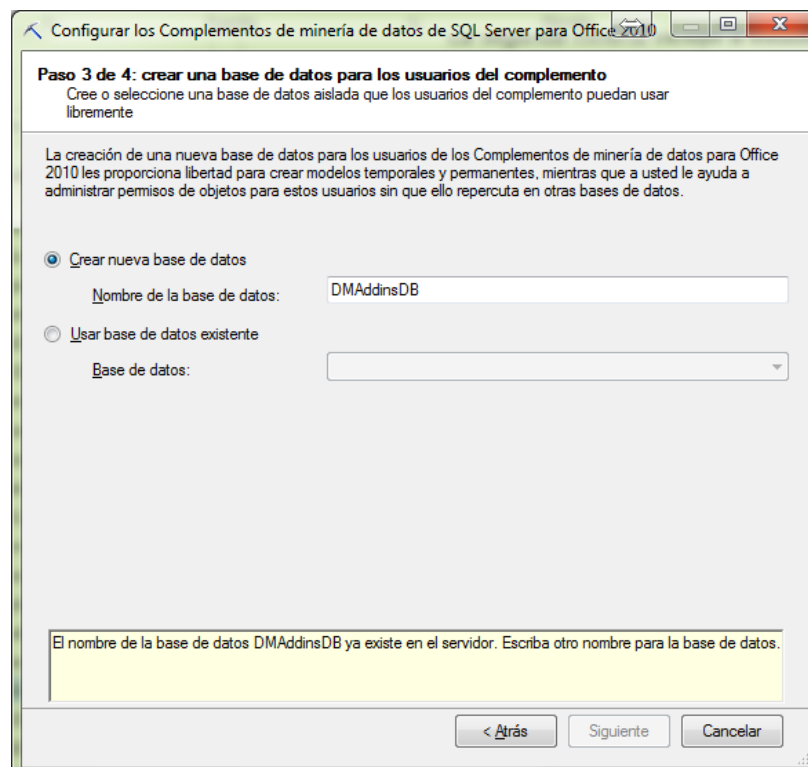


Figura 42. Configuración del servidor SQL para minería de datos con *Office*

En *Microsoft Excel* se utiliza datos desde otras fuentes y se selecciona la tabla que servirá para el análisis de los estudiantes, previamente generada mediante un proceso de carga. En la Figura 43 se observa los datos recuperados desde la tabla de análisis de estudiantes y listo para su uso con el complemento de minería de datos.

	F	G	H	I	J	K	L
	NOMBRE_ESTUDIANTE	MODALIDAD_ESTUDIANTE	TIPO_BECA_ESTUDIANTE	SEXO_ESTUDIANTE	ESTADO_CIVIL_ESTUDIANTE	FECHA_NACIMIENTO_ESTUDIANTE	PAIS_ORIGEN_ESTUDIANTE
1	DIANA VICTORIA	PRESENCIAL	SIN DATO	F	SOLTERO	27/01/1987 0:00	ECUADOR
2	ADITA LOURDES	PRESENCIAL	SIN DATO	F	SOLTERO	19/03/1974 0:00	ECUADOR
3	JORGE ENRIQUE	PRESENCIAL	SIN DATO	M	CASADO	28/07/1970 0:00	ECUADOR
4	LILIANA JACKELINE	PRESENCIAL	SIN DATO	F	VIUDO	12/09/1965 0:00	ECUADOR
5	DENNISE LIZBETH	PRESENCIAL	SIN DATO	F	SOLTERO	23/03/1986 0:00	ECUADOR
6	EDGAR VINICIO	PRESENCIAL	SIN DATO	M	SOLTERO	08/08/1985 0:00	ECUADOR

Figura 43. Pantalla del *Add-In* de minería de datos en *Microsoft Excel*

Para la primera minería se utilizó los algoritmos de clasificación permitiendo generar varios modelos, y extraer patrones que predicen el valor de una columna en este caso el estado de graduación del estudiante. La Figura 44 muestra el asistente de clasificación de *Microsoft Excel* la selección en la columna que se va a analizar, con las columnas que funcionaran cómo entradas para formar el modelo.

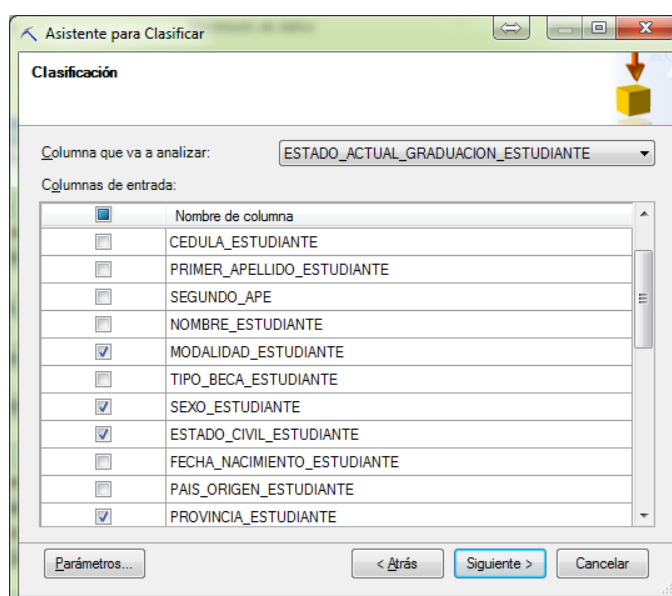


Figura 44. Asistente para crear el modelo y clasificar

En la Figura 45 se observa los diferentes algoritmos que utiliza *Microsoft Excel* en los métodos de clasificación.

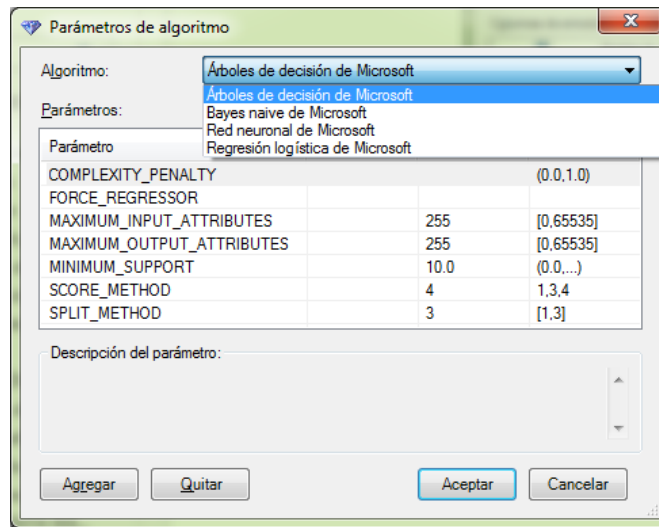


Figura 45. Parámetros de los algoritmos de clasificación

La Figura 46 muestra el asistente que sirve para dividir los datos del conjunto de entrenamiento y de pruebas, utilizadas para crear el modelo de clasificación.

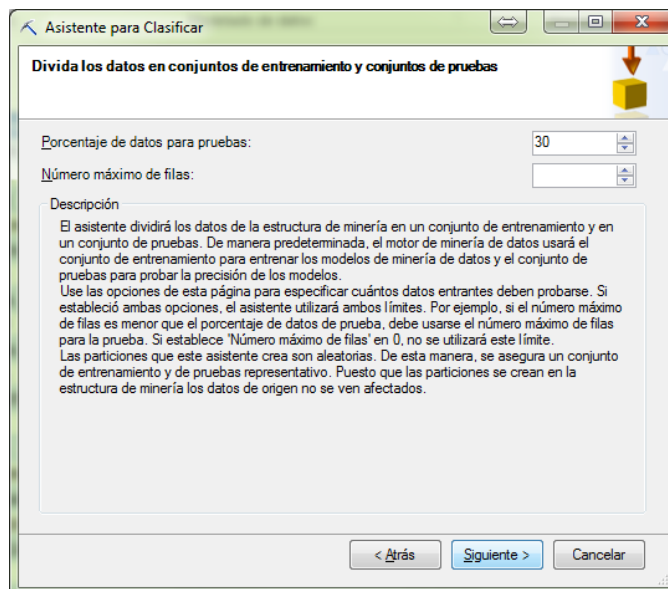


Figura 46. División de los datos de entrenamiento y conjunto de pruebas

Para el primer modelo se utilizó el algoritmo árboles de decisión de *Microsoft* como se muestra en la Figura 47, se puede ir expandiendo y analizando cada

una de sus ramificaciones, con lo que se pudo concluir que la facultad de Ciencias de la Salud tiene un porcentaje de graduación de un 99% de todos los estudiantes que terminan todas las asignaturas de estudio.

La mayor probabilidad de estudiantes que no se gradúan de acuerdo al modelo son los de la facultad de estudiantes de Educación a Distancia, su carrera sea de Administración de Empresas y su colegio de origen sea fisco misional.

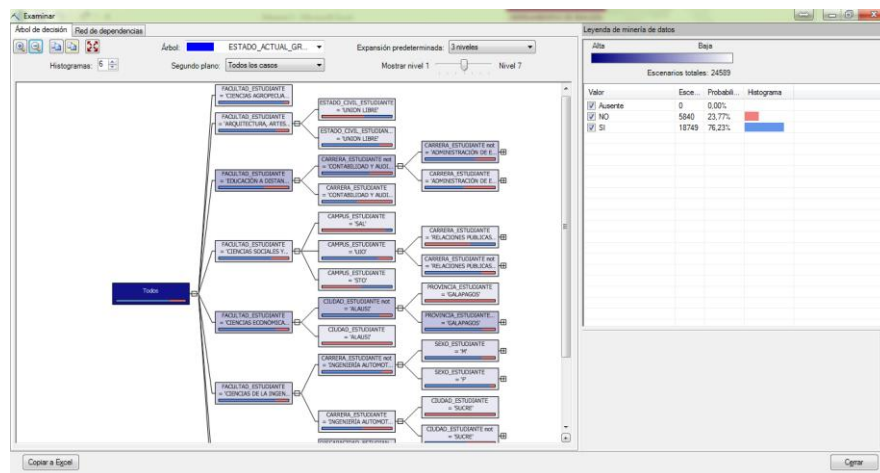


Figura 47. Pantalla árbol de decisión en estado de graduación

Para poder evaluar entre diferentes análisis predictivos es necesario agregar un nuevo modelo a la estructura creada anteriormente con otro algoritmo diferente. La Figura 48 muestra el asistente para seleccionar otro algoritmo y crear otro modelo.

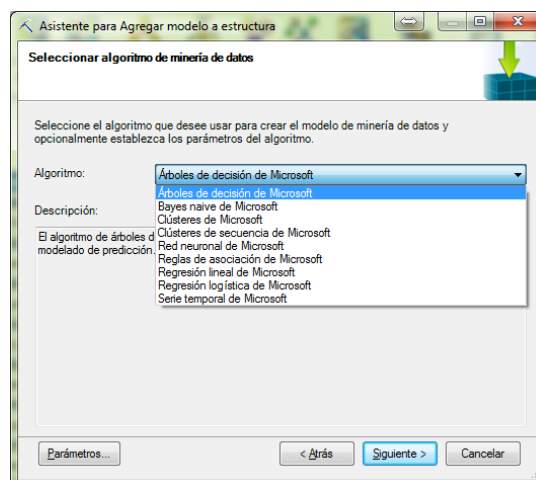


Figura 48. Asistente para agregar un modelo a la estructura

Para el siguiente modelo se seleccionó el algoritmo de *Bayes naive* de *Microsoft*. En la Figura 49 se observa el resultado del modelo. Se analiza que la mejor distribución de los estudiantes que se gradúan después de terminar todas las asignaturas de estudio es principalmente del campus de Quito, de la carrera de Ingeniería en Comercio Exterior, de estado civil soltero, de la facultad de Ciencias Económicas, de la modalidad presencial, de un colegio particular.

Los estudiantes que no se gradúan tienen una incidencia mayor en la carrera de Administración de Empresas, de estado civil soltero, de la modalidad presencial, y su colegio de origen sea fiscal.

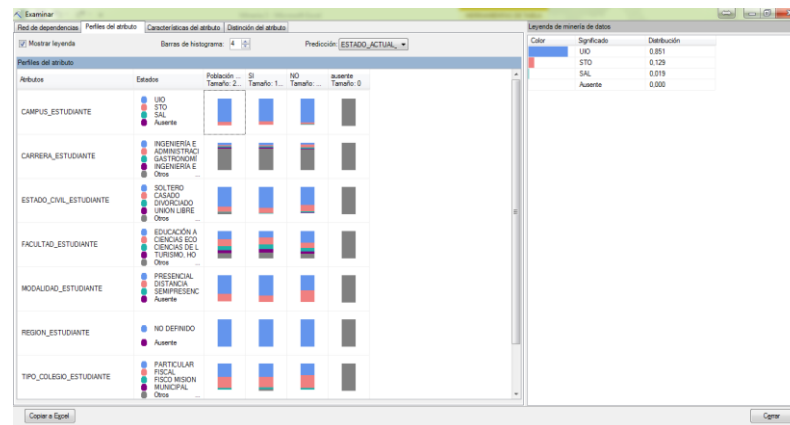


Figura 49. Pantalla de perfiles de atributos con *Bayes Naive* de *Microsoft*

Para evaluar los algoritmos de clasificación se realizó los gráficos de precisión de cada algoritmo como se muestra en la Figura 50, llegando a determinar que para los modelos, el algoritmo de árboles de decisión es el más cercano al modelo ideal.

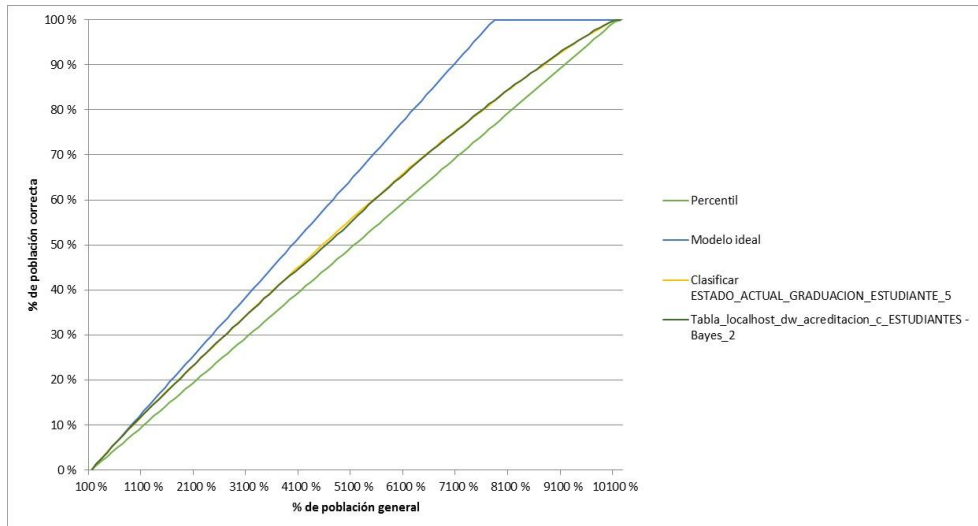


Figura 50. Gráfico de precisión de los algoritmos

Para el segundo experimento de minería de datos se realizó un proceso previo de filtrar a los estudiantes de la carrera de Ingeniería Informática que terminaron las asignaturas de estudio y se gradúan. La Figura 51 muestra a los alumnos que terminaron todas su malla curricular y se encuentran en un estado de “SI” para los graduados, un estado de “NO” para los que continúan sin graduarse. En este análisis se encontró que los estudiantes de estado civil diferente de soltero tienen un mayor índice de no graduarse, los estudiantes que su estado civil es soltero y de colegios particulares tiene una mayor probabilidad de cambiar su estado a graduados.

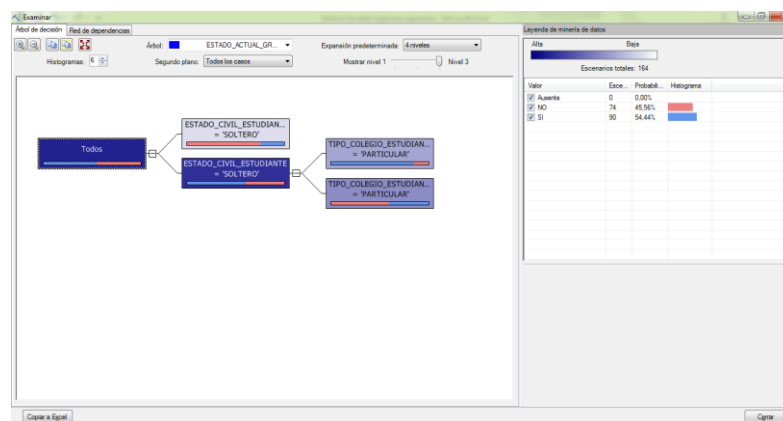


Figura 51. Pantalla de árboles de decisión de los estudiantes de Ingeniería Informática

En la Figura 52 muestra los resultados del proceso del algoritmo de *Bayes Naive* de *Microsoft*, se observa que los estudiantes que se gradúan son los de colegio particulares, ninguna discapacidad, estado civil soltero, modalidad presencial, provincia de Pichincha, sexo masculino.

Los estudiantes que terminan todas sus asignaturas de estudio y no se gradúan son principalmente de los colegios fiscales, provincia Pichincha y de estado civil soltero.

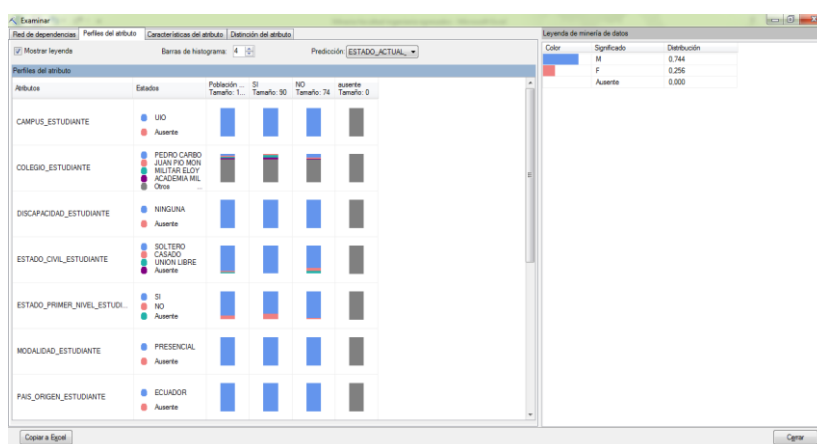


Figura 52. Pantalla de perfiles del atributo de los estudiantes de Ingeniería Informática

La Figura 53 muestra el gráfico de precisión del modelo, se observa que el algoritmo de *Bayes Naive* de *Microsoft* es el que mejor se aproxima al modelo ideal de la información examinada.

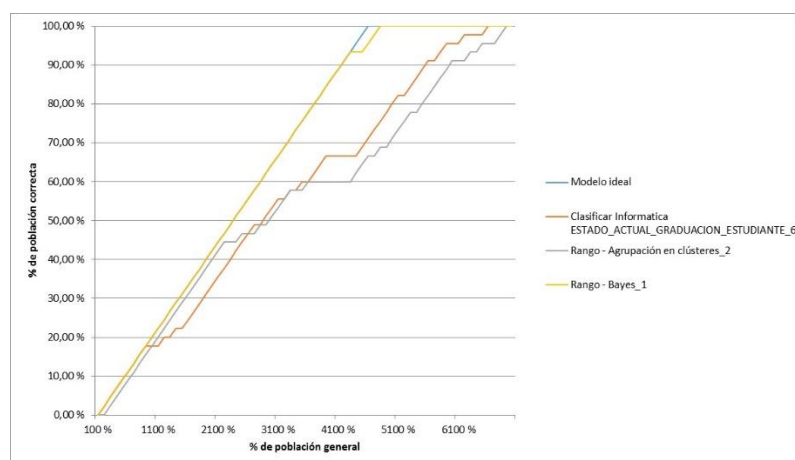


Figura 53. Gráfico de precisión del modelo de estudiantes de Ingeniería Informática

5. CONCLUSIONES Y RECOMENDACIONES

5. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- El proceso de construcción de un *data warehouse* es muy extenso debido al pre procesamiento y limpieza que se tiene que dar a los datos de acuerdo a los requisitos del negocio por lo que fue necesario escoger una metodología que facilite en la planificación por etapas del proyecto y permita un mantenimiento futuro.
- El *data warehouse* no cumplió con la finalidad de obtener datos directamente de la base operacional de la institución debido a la falta de documentación de las tablas y problemas de validación en los procesos previos de cada departamento que impiden que los datos a obtenerse sean en tiempo real.
- Un *data warehouse* construido con herramientas de inteligencia de negocios tiene muchos usos para poder analizar información dependiendo del conocimiento que se desee obtener mediante *software* de análisis apropiados.
- El *data warehouse* ayudó a determinar varios errores que se encuentran en las bases de datos operacionales de la institución lo que actualmente dificulta obtener una información precisa en los sistemas institucionales para tomar decisiones adecuadas.
- La generación de cubos multidimensionales apoyado con reportes reflejó de una manera más amplia el estado de la institución, y en el cual se podrá investigar de una manera más detallada la información en el *data warehouse* para la formulación de nuevas estrategias.
- Mediante la minería de datos se identificó tendencias desconocidas sobre los estudiantes las cuales ayudan a mejorar la gestión académica a través de la generación de estrategias que en función del tiempo aporten al mejoramiento de los indicadores de la universidad.

5.2. RECOMENDACIONES

- Los archivos modelos de *Microsoft Excel* que son enviados por el ITIC que funcionan como origen de datos necesitan que sean verificados previamente y mantengan el formato correspondiente para no tener problemas en el proceso ETL.
- Los errores encontrados en los datos necesitan corregirse directamente en la base operacional mediante una auditoria informática determinando la razón con sus respectivos correctivos en los procesos previos para mejorar el funcionamiento de los sistemas institucionales y mantener una información unificada.
- Para tener un manejo preciso de la información y con el objetivo de consolidar la misma se recomienda codificar cada uno de los departamentos dentro de las distintas sedes de la universidad aportando a la mejora de la cultura organizacional de la institución.
- Para una implementación futura del *data warehouse* se necesita trabajar en conjunto con el ITIC y con los *stakeholder* de cada departamento con el fin de conformar indicadores representativos que sean objeto de medición para la institución.
- Para un correcto desempeño de las transformaciones mediante ETL se recomienda madurar los procesos previos de recolección de información, para garantizar una estructura institucional con procedimientos estables a fin de sustentar una fuente de origen de calidad para la integración de datos.
- Se recomienda la utilización de herramientas con análisis estadístico y dinámico como *dashboard* que faciliten la interpretación de gráficos analíticos con una amplia visualización de estado institucional, o con procesos de minería de datos que permitan encontrar otros indicadores valiosos y generen proyecciones que beneficien a la institución.

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Abella, R., Cóppola, L., & Olava, D. (2000). *Sistema data warehousing Carga y Control de Calidad*. Universidad de la República. Retrieved from http://www.fing.edu.uy/~ruggia/T5s/T5DWBede98ETL_InfoPrincipal.pdf
- Avilés, A. (2014). Diseño de un Modelo de Gestión del Conocimiento para desarrollar Competencias Investigativas en la educación media y superior en Guayaquil. *Uma ética Para Quantos?*, XXXIII(2), 81–87. <http://doi.org/10.1007/s13398-014-0173-7.2>
- Ballesteros Román, A., Sánchez-Guzmán, D., & García Salcedo, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patroness de aprendizaje sobre un contexto educativo. *Latin-American Journal of Physics Education*, 7(4), 662–668. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=94988767&lang=es&site=ehost-live>
- Benavides, C., & Quintana, C. (2005). Proceso y sistemas organizativos para la gestión del conocimiento El papel de la calidad total. *Boletín Económico de ICE*, 2838, 37–52. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=2051365>
- Boronat, M., Villar, A., & Puig, A. (2007). Sistemas De Gestión De La Calidad Total , Gestión Del Conocimiento Y De La I + D Y Desempeño Organizativo, 1–14.
- Calva, J. (2013). Una aproximación a lo que son las necesidades de información, 1–4. Retrieved from http://bvs.sld.cu/revistas/mciego/alfin_2012/alfin_folder/2012_Unidad3/Bibliograf?aLecturasB?sicasycomplementarias/Lectura_basica_1.Una_aproximacion_a_lo_que_son_las_necesidades_de_informacion.pdf
- Casanova, C. (2009). El rol de los sistemas de información en la gestión del

conocimiento en las empresas, 40 –52. Retrieved from http://www.petrotecnica.com.ar/junio09/el_rol.pdf

Connolly, T., & Begg, C. (2005). *Database Systems* (Fourth). Englad: British Library Cataloguing.

Contreras, E. (2010). Gestión del Conocimiento: Del tácito al explícito 20 años después. *Trend Management. Edición Especial Mayo 2010*, 94–100. Retrieved from www.dii.uchile.cl/wp-content/uploads/2011/06/UCH-Contreras.pdf

Duque, A. F. (2010). *Implementación de un Datawarehouse para el Instituto Geográfico Militar*. Escuela Politécnica del Ejército. Retrieved from <http://repositorio.espe.edu.ec/handle/21000/336>

Figueroa, M. (2006). *Minería de Datos Aplicada a Credit Scoring*. Universidad San Francisco de Quito. Retrieved from <http://repositorio.usfq.edu.ec/bitstream/23000/547/1/82579.pdf>

Funes, Y., & Hernández, C. (2001). Medición del valor del capital intelectual. *Contaduría Y Administración*, (203), 45–58. Retrieved from <http://www.revistas.unam.mx/index.php/rca/article/view/4527>

García, H. Á. (2010). Minería de Datos en la Educación, 2 –8. Retrieved from <http://www.it.uc3m.es/jvillena/irc/practicas/10-11/08mem.pdf>

Gomez, A. (2010). INTELIGENCIA DE NEGOCIOS: ESTADO DEL ARTE BUSINESS INTELLIGENCE: STATE OF THE ART. *Scientia Et Technica*, XVI(44), 321–326. Retrieved from <http://www.redalyc.org/articulo.oa?id=84917316060>

Gonzalez Millan, J. (2009). *El Capital Intelectual y sus indicadores en el sector Industrial*. Universidad Pedagógica y Tecnológica de Colombia.

Guanotásig, A. (2013). *Desarrollo de una solución datawarehouse para el sistema de cartera en la PyME “La hipotecaria.”* Universidad Tecnológica Equinoccial, Quito.

- Hernández, K. (2003). *Propuesta para el tratamiento contable - financiero del Capital Intelectual en México*. Universidad de las Américas Puebla. Retrieved from http://catarina.udlap.mx/u_dl_a/tales/documentos/lcp/morales_o_ls/
- Ilbay, E. (2009). *Propuesta metodológica para aplicar business intelligence caso practico Cohervi S.A*. Escuela Superior Politécnica de Chimborazo. Retrieved from <http://dspace.esPOCH.edu.ec/bitstream/123456789/57/1/18T00388.pdf>
- Kaplan, R., & Norton, D. (2005). El cuadro de mando integral, 1–7. Retrieved from http://www.factorhumana.org/attachments_secure/article/8312/UC_QCI_cas.pdf
- Langa, M. E., Carbonell, M. F., Vicedo, J. C., & Miquel, V. (2005). El capital intelectual como activo estratégico en la sociedad del conocimiento . Una aproximación a su valoración mediante la aplicación del modelo Technology Broker en una empresa del sector cosméticos. *IX Congreso de Ingeniería de Organización*, 1 –8. Retrieved from <http://www.adingor.es/Documentacion/CIO/cio2005/items/ponencias/16.pdf>
- León, M., & Mercader, R. (2002). LOS PROCESOS DE CREACIÓN DEL CONOCIMIENTO. *XVI Congreso Nacional de La Asociación Española de Dirección Y Economía de La Empresa*, 1–16. Retrieved from <http://www.upct.es/~economia/PUBLI-INO/LOS PROCESOS DE CREACION DEL CONOCIMIENTO- EL APRENDIZA.pdf>
- Lidia, N., Aguiar, I., & Saá, P. (2008). La incidencia de la adquisición y generación de conocimiento en la innovación, *17*, 57–77.
- Martínez, J. (2008). *Diseño del catálogo de metadatos para la automatización del proceso de carga de un data webhouse Agradecimientos*. Universidad Politécnica de Madrid. Retrieved from

http://oa.upm.es/1024/1/PFC_JUAN_LUIS_MARTINEZ_ALBURNE.pdf

Monsalve, J., Aponte, F., & Hoyos, J. (2013). Aplicación de minería de datos educativos a procesos B-Learning. *Conferencias LACLO*, 4(1). Retrieved from <http://www.laclo.org/papers/index.php/laclo/article/view/75>

Moreno, M., Quintales, L., Garcia, F., & Polo, J. (2008). Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil. *liis.Org*, 14. Retrieved from http://www.iiis.org/CDs2010/CD2010CSC/CISCI_2010/PapersPdf/CA156FK.pdf

Nader, J. (2011). *Sistema de Apoyo Gerencial Universitario*. Instituto Tecnológico Buenos Aires. Retrieved from <http://www2.itba.edu.ar/archivos/secciones/nader-tesisdemagister.pdf>

Oracle, C. (2009). ¿Qué es Inteligencia de Negocios?, 1–6. Retrieved from http://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf

Pávez Salazar, A. A. (2000). *Modelo de implantación de gestión del conocimiento y tecnologías de información para la generación de ventajas competitivas*. Valparaíso: Universidad Técnica Federico Santa María. Universidad Técnica Federico Santa María. Retrieved from http://sgpwe.izt.uam.mx/files/users/uami/sppc/GC_Literatura/Pavez_Alejandro_2001_GC_y_TI_para_crear_Ventajas_competitivas.pdf

Pazmiño, S., & Suasnavas, D. (2012). *Diseño, desarrollo e implementación de un sistema para la gestión de clientes y reservas al centro cultural del Instituto Geografico Militar*. Universidad Central del Ecuador. Retrieved from <http://www.dspace.uce.edu.ec/bitstream/25000/3014/1/T-UCE-0011-123.pdf>

Pignani, J. M. (2011). *Sistemas Expertos (Expert System) Orientación I: Informática aplicada a la ingeniería de procesos 1*. Universidad Tecnológica Nacional, Facultad Regional de Rosario, (Orientación I).

- Retrieved from
http://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientador_a1/monograis/pignani-sistemasexpertos.pdf
- Ril, V., Rodríguez, R., Piñero, P. Y., & Martínez, H. A. (2013). Knowledge discovery in lessons learned documented of closure processes in IT projects. *Revista Cubana de Ciencias Informáticas*, 7(3), 45–57. Retrieved from
http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2227-18992013000300005&lng=en&nrm=iso&tlng=en
- Rivadera, G. R. (2010). La metodología de Kimball para el diseño de almacenes de datos (Data warehouses), 56–71. Retrieved from
<https://padlet-uploads.storage.googleapis.com/70859366/07c12e2c6b993f9774da2a14c585e26c499ab6a8/d23a819e370ac2b4e3e713dd4d59de2a.pdf>
- Rojas, I. (2009). Data Warehouse, 1–37. Retrieved from
<http://exa.unne.edu.ar/informatica/SO/MonoAdsDiseno.pdf>
- Rotaeché, E. (2007). Business Intelligence, 7. Retrieved from
<https://churriwifi.files.wordpress.com/2009/11/business-intelligence-ibermatica.pdf>
- Sánchez, M. (2005). Breve inventario de los modelos para la gestión del conocimiento en las organizaciones. *Breve Inventario de Los Modelos Para La Gestión Del Conocimiento En Las Organizaciones*, 13(6), 18. Retrieved from http://bvs.sld.cu/revistas/aci/vol13_6_05/aci060605.pdf
- Santamaría, W. (2010). *Modelo de detección de fraude basado en el descubrimiento simbólico de reglas de clasificación extraídas de una red neuronal*. Universidad Nacional de Colombia. Retrieved from
<http://www.bdigital.unal.edu.co/3086/1/299742.2010.pdf>
- Silva, M. (2007). MINERIA DE DATOS Y Descubrimiento del Conocimiento. *Universidad Nacional Del Noreste, Facultad de Ciencias Exactas Y*

Naturales. Retrieved from
http://exa.unne.edu.ar/informatica/SO/Mineria_de_Datos_y_KDD.pdf

Valencia Arcos, J. del C., & Guevara Lenis, J. E. (2007). *Data warehouse para el análisis académico de la Escuela Politécnica Nacional*. Escuela Politecnica Nacional. Retrieved from
<http://bibdigital.epn.edu.ec/handle/15000/445>

Velasco, C., & Quintana, C. (2003). *GESTIÓN DEL CONOCIMIENTO Y CALIDAD TOTAL*. España.

Ydirín, M. (2004). *Construcción de un Data Warehouse de datos del medio ambiente para la toma de decisiones: aplicación a los datos hidrológicos*. Universidad de las Américas Puebla. Retrieved from
http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/ydirin_p_mm/

Zapata, L. (2004). *Los Determinantes de la Generación y la Transferencia del Conocimiento en Pequeñas y Medianas Empresas del Sector de las Tecnologías de la Información de Barcelona*. Universidad Autonoma de Barcelona. Retrieved from
<http://www.tdx.cat/bitstream/handle/10803/3955/lezc1de1.pdf?sequence=1>

ANEXOS

Anexo 1. Tabla de hechos porcentaje meta

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactPorcentajeCumplimientoMeta	Inversion_Id	RptTotalU ser	C_adm
FactPorcentajeCumplimientoMeta	Kpi_Id	RptTotalU ser	KPI
FactPorcentajeCumplimientoMeta	Unidades_Plani_Id	RptTotalU ser	C_UNIT
FactPorcentajeCumplimientoMeta	Codigo_Estrategico_Id	RptTotalU ser	OBJETIVOS TÁCTICOS
FactPorcentajeCumplimientoMeta	Tiempo_Plazo_Id	RptTotalU ser	Plazo
FactPorcentajeCumplimientoMeta	Codigo_Meta_Id	RptTotalU ser	CMM
FactPorcentajeCumplimientoMeta	Objetivo_Unidad_Id	RptTotalU ser	Objetivos de la Unidad
FactPorcentajeCumplimientoMeta	Pm_Id	RptTotalU ser	PM
FactPorcentajeCumplimientoMeta	Meta_Inversion_Gasto_Id	RptTotalU ser	Inversión/Gas to
FactPorcentajeCumplimientoMeta	Retroalimentacion_Id	RptTotalU ser	Fecha Retroalimentación / Retroalimentación
FactPorcentajeCumplimientoMeta	Logros_Alcanzados_Id	RptTotalU ser	Fecha Actualización Unidad / Observaciones / Lista de evidencias / Logros alcanzados

Anexo 1. Tabla de hechos porcentaje meta continuación

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactPorcentajeCumplimientoMeta	Objetivo_Institucional_Id	RptTotalU ser	Objetivo Institucio nal
FactPorcentajeCumplimientoMeta	Indicador_Logro_Id	RptTotalU ser	Indicador de Logro de la Meta
FactPorcentajeCumplimientoMeta	Usuario_Unidad_Id	RptTotalU ser	Usuario Unidad
FactPorcentajeCumplimientoMeta	Metas_Cumplirse_Id	RptTotalU ser	Plazo / Meta a Cumplirs e
FactPorcentajeCumplimientoMeta	Porcentaje_Cumplimiento _Meta	RptTotalU ser	% Cumplido

Anexo 2. Dimensión Inversión

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCodigoInversionGasto	Codigo_Inversion	Codigos_Invers gasto_ADM	Código
DimCodigoInversionGasto	Significado_Invers ion	Codigos_Invers gasto_ADM	Significado

Anexo 3. Dimensión Código Kpi

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCodigoKpi	Codigo_Kpi	Codigos_KPI	Código
DimCodigoKpi	Significado_Codig o	Codigos_KPI	Significado

Anexo 4. Dimensión Unidades

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCodigoKpi	Codigo_Kpi	Unidades_2	Código
DimCodigoKpi	Significado_Codig o	Unidades_2	Significado

Anexo 5. Dimensión Objetivo Unidad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimObjetivosEstrategicos Unidad	Nombre_Objetivo	CodigosEstrate gias	OBJETIVO ESTRATÉ GICO
DimObjetivosEstrategicos Unidad	Estrategias_Defini ción	CodigosEstrate gias	ESTRATE GIAS
DimObjetivosEstrategicos Unidad	Objetivos_Tactico s	CodigosEstrate gias	OBJETIVO S TÁCTICOS

Anexo 6. Dimensión Tiempo

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimTiempo	Anio_Tiempo	Tiempo 1900-2020	Anio
DimTiempo	Trimestre_Tiempo	Tiempo 1900-2020	Trimestre
DimTiempo	Semestre_Tiempo	Tiempo 1900-2020	Semestre
DimTiempo	Mes_Numero_Tiempo	Tiempo 1900-2020	MesNumero
DimTiempo	Mes_Letra_Tiempo	Tiempo 1900-2020	MesLetra
DimTiempo	Semana_Mes_Tiempo	Tiempo 1900-2020	SemanaMes
DimTiempo	Dia_Semana_Numero	Tiempo 1900-2020	DiaSemanaNumero
DimTiempo	Dia_Semana_Letra	Tiempo 1900-2020	DiaSemanaLetra
DimTiempo	Dia_Anio_Mes_Numero	Tiempo 1900-2020	Dia_anio_mes_numero

Anexo 7. Dimensión Código Meta Mes

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCodigoMeta	Codigo_Meta_Mes	RptTotalUser	CMM
DimCodigoMeta	Codigo_Meta	RptTotalUser	Cód# Meta

Anexo 8. Dimensión Meta Cumplirse

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaCumplirse	Meta_Cumplir	RptTotalUser	Metas a Cumplirse
DimMetaCumplirse	Plazo_Meta_Cumplirse	RptTotalUser	Plazo
DimMetaCumplirse	Peso_Meta_Cumplirse	RptTotalUser	Peso Meta

Anexo 9. Dimensión Meta Indicador Logro

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaIndicadorLogro	Indicador_Logro	RptTotalUser	Indicador de Logro de la Meta

Anexo 10. Dimensión Inversión gasto

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaInversionGasto	Inversion_Gasto	RptTotalUser	Inversión/Gasto

Anexo 11. Dimensión Logros Alcanzados

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaLogrosAlcanzados	Logros_Alcanzados	RptTotalUser	Logros alcanzados
DimMetaLogrosAlcanzados	Observaciones	RptTotalUser	Observaciones
DimMetaLogrosAlcanzados	Lista_Evidencias	RptTotalUser	Lista de Evidencias
DimMetaLogrosAlcanzados	Fecha_Actualización_Unidad	RptTotalUser	Fecha Actualización Unidad

Anexo 12. Dimensión Objetivo Institucional

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaObjetivoInstitucional	Objetivo_Institucional	RptTotalUser	Objetivo Institucional
DimMetaObjetivoInstitucional	Numero_Objetivo_Institucional	RptTotalUser	OI

Anexo 13. Dimensión Objetivo Unidad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaObjetivoUnidad	Objetivo_Unidad	RptTotalUser	Objetivos de la Unidad
DimMetaObjetivoUnidad	Peso_Objetivo_Unidad	RptTotalUser	Peso Obj# Unidad

Anexo 14. Dimensión Meta Plan de Mejora

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaPm	PM	RptTotalUser	PM
DimMetaPm	Responsable_Pm	RptTotalUser	Responsable

Anexo 15. Dimensión Meta Retroalimentación

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaRetroalimentacion	Retroalimentacion	RptTotalUser	Retroalimentación
DimMetaRetroalimentacion	Fecha_Retroalimentacion	RptTotalUser	Fecha Retroalimentación

Anexo 16. Dimensión Meta Usuario Unidad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMetaUsuarioUnidad	Usuario_Unidad	RptTotalUser	Usuario Unidad

Anexo 17. Dimensión Biblioteca

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimBiblioteca	Nombre_Biblioteca	Hoja1	NOMBRE_BIBLIOTECA

Anexo 18. Dimensión Campus

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCampus	Nombre_Campus	Hoja1	NOMBRE_CAMPUS

Anexo 19. Dimensión Aulas

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimAulas	Codigo_Aulas	Sheet1	CODIGO_AULAS
DimAulas	Facilidad_Enseñanza	Sheet1	AULAS_CON_FACILIDADES_PARA_ENSEÑANZA

Anexo 20. Dimensión Presupuesto

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimPresupuestoUniversitario	Nombre_Cuentas	Sheet1	NOMBRE_CUENTA

Anexo 21. Dimensión Materias

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimMateria	Codigo_Materia	Sheet1	cod_materia
DimMateria	Nombre_Materia	Sheet1	materia

Anexo 22. Dimensión Carreras

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCarrera	Codigo_Carrera	Sheet1	cod_carrera
DimCarrera	Nombre_Carrera	Sheet1	carrera

Anexo 23. Dimensión Facultad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimFacultad	Nombre_Facultad	Sheet1	facultad

Anexo 24. Dimensión Docente

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimDocente	Cedula_Docente	UIO / SAL /STO	CEDULA
DimDocente	Primer_Apellido_Docente	UIO / SAL /STO	PRIMER_APELLIDO
DimDocente	Segundo_Apellido_Docente	UIO / SAL /STO	SEGUNDO_APELLIDO
DimDocente	Nombre_Docente	UIO / SAL /STO	NOMBRES
DimDocente	Nacionalidad_Docente	UIO / SAL /STO	NACIONALIDAD
DimDocente	Fecha_Nacimiento	UIO / SAL /STO	FECHA_NACIMIENTO
DimDocente	Relacion_Trabajos	UIO / SAL /STO	RELACION_TRABAJOS
DimDocente	Unidad_Academica	UIO / SAL /STO	UNIDAD_ACADEMICA
DimDocente	Movilidad_les_Extranjera	UIO / SAL /STO	MOVILIDAD_IES_EXTRANJERAS
DimDocente	Movilidad_les_Nacional	UIO / SAL /STO	MOVILIDAD_IES NACIONALES

Anexo 25. Dimensión Tiempo dedicación

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimTiempoDedicacion	Tiempo_Dedicacion	UIO / SAL /STO	TIEMPO_DEDICACION

Anexo 26. Dimensión Período Sabático

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimPeriodoSabatico	Periodo_Sabatico_Docente	UIO / SAL /STO	PERIODO_SABATICO
DimPeriodoSabatico	Fecha_Inicio_Periodo_Sabatico	UIO / SAL /STO	FECHA_INICIO_SABATICO
DimPeriodoSabatico	Fecha_Fin_Periodo_Sabatico	UIO / SAL /STO	FECHA_FIN_SABATICO

Anexo 27. Dimensión Ingreso Concurso

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimIngresoConcurso	Ingreso_Concurso	UIO / SAL /STO	INGRESO_CONCURSO

Anexo 28. Dimensión Sub área Último Título Cuarto Nivel

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimSubareaUltimoTituloCuarto	Subarea_Ultimo_Titulo_Cuarto	UIO / SAL /STO	SUBAREA _ULTIMO_ TITULO_C UARTO_NI VEL

Anexo 29. Dimensión Período

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimPeriodo	Periodo_Real_Detalle	Sheet1	periodo
DimPeriodo	Anio_Periodo	Sheet1	anio

Anexo 30. Dimensión Cursando Estudios

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCursandoEstudios	Cursando_Estudios	UIO / SAL /STO	CURSAND O_ESTUDI OS
DimCursandoEstudios	Financiamiento_Estudios	UIO / SAL /STO	FINANCIA MIENTO_ ESTUDIO S_PHD

Anexo 31. Dimensión Cargo Autoridad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCargoAutoridad	Cargo_Autoridad	UIO / SAL / STO	CARGO_A UTORIDA D

Anexo 32. Dimensión Último Título Cuarto Nivel

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimUltimoTituloCuartoNivel	Nivel_Ultimo_Titulo_Cuarto_Nivel	UIO / SAL / STO	NIVEL_ ULTIMO_ TITULO_C UARTO_NI VEL
DimUltimoTituloCuartoNivel	Ultimo_Titulo_Cuarto_Nivel	UIO / SAL / STO	ULTIMO_ TITULO_C UARTO_NI VEL
DimUltimoTituloCuartoNivel	Fecha_Ultimo_Titulo_Cuarto_Nivel	UIO / SAL / STO	FECHA_U LTIMO_TI TULO_CU ARTO_NIV EL
DimUltimoTituloCuartoNivel	Universidad_Ultimo_Titulo_Cuarto_Nivel	UIO / SAL / STO	UNIVERSI DAD_ ULTIMO_ TITULO_C UARTO_NI VEL
DimUltimoTituloCuartoNivel	Pais_Ultimo_Titulo_Cuarto_Nivel	UIO / SAL / STO	PAIS_ ULTIMO_ TITULO_C UARTO_NI VEL

Anexo 33. Dimensión Categoría Personal Académico

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimCategoriaPersonalAcademico	Categoria_Personal_Academico	UIO / SAL / STO	CATEGORIA_PERSONAL_ACADEMICO

Anexo 34. Dimensión Último Título Tercer Nivel

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimUltimoTituloTercerNivel	Nivel_Ultimo_Titulo_Tercer_Nivel	UIO / SAL / STO	NIVEL_ULTIMO_TITULO_TERCER_NIVEL
DimUltimoTituloTercerNivel	Ultimo_Titulo_Tercer_Nivel	UIO / SAL / STO	ULTIMO_TITULO_TERCER_NIVEL

Anexo 35. Dimensión Sexo

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimSexo	Nombre_Sexo	UIO / SAL / STO / Sheet1	sexo

Anexo 36. Dimensión Modalidad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimModalidad	Nombre_Modalidad	UIO / SAL / STO / Sheet1	modalidad

Anexo 37. Dimensión Colegio

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimColegio	Tipo_Colegio	Sheet1	colegio tipo
DimColegio	Nombre_Colegio	Sheet1	colegio nombre

Anexo 38. Dimensión Discapacidad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimDiscapacidad	Nombre_Discapacidad	Sheet1	discapacidad

Anexo 39. Dimensión Estado Civil

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimEstadoCivil	Nombre_Estado_Civil	Sheet1	estado_civil

Anexo 40. Dimensión Localidad

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimLocalidadOrigen	Pais_Origen	Sheet1	colegio tipo
DimLocalidadOrigen	Provincia_Origen	Sheet1	colegio nombre
DimLocalidadOrigen	Ciudad_Origen	Sheet1	ciudad

Anexo 41. Dimensión Estudiantes Primer Nivel

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimEstudiantesPrimerNivel	Cedula_Estudiantes_Primer	Sheet1	cedula
DimEstudiantesPrimerNivel	Primer_Apellido_Estudiante_Primer	Sheet1	primer_apellido
DimEstudiantesPrimerNivel	Segundo_Apellido_Estudiante_Primer	Sheet1	segundo_apellido
DimEstudiantesPrimerNivel	Nombre_Estudiante_Primer	Sheet1	nombre
DimEstudiantesPrimerNivel	Telefono_Estudiante_Primer	Sheet1	telefono1
DimEstudiantesPrimerNivel	Email_Estudiante_Primer	Sheet1	email_personal
DimEstudiantesPrimerNivel	Periodo_Estudiante_Primer	Sheet1	periodo

Anexo 42. Dimensión Estudiante Posgrado

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimEstudiantesPos	Cedula_Estudiantes_Pos	Sheet1	cedula
DimEstudiantesPos	Primer_Apellido_Estudiante_Pos	Sheet1	primer_apellido
DimEstudiantesPos	Segundo_Apellido_Estudiante_Pos	Sheet1	segundo_apellido
DimEstudiantesPos	Nombre_Estudiante_Primer	Sheet1	nombre
DimEstudiantesPos	Fecha_Egresamiento	Sheet1	fecha_egresamiento

Anexo 43. Dimensión Publicación Artículos

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimPublicacionArticulo	Nombre_Publicación	Libros	NOMBRE_PUBLICACION
DimPublicacionArticulo	Tipo_Publicacion	Libros	TIPO_PUBLICACION
DimPublicacionArticulo	Participación_Publicacion	Libros	PARTICIPACION_PUBLICACION
DimPublicacionArticulo	Numero_Isbn	Libros	NUMERO_ISBN
DimPublicacionArticulo	Revision_Publicacion_Pares	Libros	REVISION_PUBLICACION_PARES
DimPublicacionArticulo	Nombre_Revista_Articulo_Publicado	Libros	NOMBRE_REVISTA_ARTICULO_PUBLICADO
DimPublicacionArticulo	Estado_Publicacion	Libros	ESTADO_PUBLICACION

Anexo 44. Dimensión Libro Capitulo

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimLibroCapitulo	Libro_Articulo_Capitulo	Libros	LIBRO_ARTICULO_CAPITULO

Anexo 45. Dimensión Estudiantes Pregrado

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimEstudiantesPre	Cedula_Estudiant e	Sheet1	cedula
DimEstudiantesPre	Primer_Apellido_E studiante	Sheet1	primer_ape llido
DimEstudiantesPre	Segundo_Apellido _Estudiante	Sheet1	segundo_a pellido
DimEstudiantesPre	Nombre_Estudiant e	Sheet1	nombre
DimEstudiantesPre	Tipo_Beca_Estudi ante	Sheet1	tipo_beca
DimEstudiantesPre	Discapacidad_Est udiante	Sheet1	discapacid ad
DimEstudiantesPre	Fecha_Convalidac ion	Sheet1	fecha_con validacion
DimEstudiantesPre	Estado_Matricula_ Actual_Estudiante s	Sheet1	estado_ma tricula_act ual
DimEstudiantesPre	Telefono_Estudian te	Sheet1	telefono1
DimEstudiantesPre	Email_Estudiante	Sheet1	email_pers onal
DimEstudiantesPre	Fecha_Nacimiento _Estudiante	Sheet1	fecha_naci miento

Anexo 46. Dimensión Revista Indexada

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimBaseDatosRevistaIndexada	Base_Datos_Revista_Indexada	Libros	BASE_DATOS_REVISTA_INDEXADA

Anexo 47. Dimensión Proyectos de Vinculación

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimProyectosVinculacion	Codigo_Proyecto_Vinculacion	Proyectos IV – V - VI	COD_VINCULACION
DimProyectosVinculacion	Nombre_Proyecto_Vinculacion	Proyectos IV – V - VI	ESTADO_PROYECTO_VINCULO
DimProyectosVinculacion	Estado_Proyecto_Vinculacion	Proyectos IV – V - VI	ESTADO_PROYECTO_VINCULO1
DimProyectosVinculacion	Director_Proyecto_Vinculacion	Proyectos IV – V - VI	DIRECTOR_PROYECTO

Anexo 48. Dimensión Proyectos de Investigación

Tabla destino	Campo destino	Tabla origen	Campo Origen
DimProyectosInvestigacion	Codigo_Proyecto_Vinculacion	Hoja1	COD_PROYECTO_INVESTIGACION
DimProyectosInvestigacion	Nombre_Proyecto_Vinculacion	Hoja1	NOMBRE_PROYECTO_INVESTIGACION
DimProyectosInvestigacion	Estado_Proyecto_Vinculacion	Hoja1	ESTADO_PROYECTO
DimProyectosInvestigacion	Director_Proyecto_Vinculacion	Hoja1	DIRECTOR_PROYECTO

Anexo 49. Tabla de Hechos Estudiantes Primer Nivel

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactEstudiantesPrimerNivel	Estudiantes_Primer_Nivel_Id	DimEstudiantesPrimerNivel	Estudiantes_Primer_Nivel_Id
FactEstudiantesPrimerNivel	Carrera_Id	DimCarrera	Carrera_Id
FactEstudiantesPrimerNivel	Facultad_Id	DimFacultad	Facultad_Id
FactEstudiantesPrimerNivel	Campus_Id	DimCampus	Campus_Id
FactEstudiantesPrimerNivel	Periodo_Id	DimPeriodo	Periodo_Id
FactEstudiantesPrimerNivel	Fecha_Admission_Tiempo_Id	DimTiempo	Tiempo_Id
FactEstudiantesPrimerNivel	Sexo_Id	DimSexo	Sexo_Id

Anexo 50. Tabla de Hechos Internet

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactInternet	Campus_Id	DimCampus	Campus_Id
FactInternet	Fecha_Reporte_Internet_Tiempo_Id	DimTiempo	Tiempo_Id

Anexo 51. Tabla de Hechos Proyectos de Vinculación

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactEstudiantesPrimerNivel	Proyectos_Vinculacion_Id	DimProyectosVinculacion	Proyectos_Vinculacion_Id
FactEstudiantesPrimerNivel	Fecha_Inicio_Vin_Tiempo_Id	DimTiempo	Tiempo_Id
FactEstudiantesPrimerNivel	Fecha_Fin_Vin_Dim_Tiempo_Id	DimTiempo	Tiempo_Id
FactEstudiantesPrimerNivel	Campus_Id	DimCampus	Campus_Id

Anexo 52. Tabla de Hechos Presupuesto Universitario

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactPresupuesto	Presupuesto_U_Id	DimPresupuest oUniversitario	Presupuest o_U_Id
FactPresupuesto	Anio_Presupuesto _Tiempo_Id	DimTiempo	Tiempo_Id
FactPresupuesto	Campus_Id	DimCampus	Campus_I d

Anexo 53. Tabla de Hechos Estudiantes Posgrado

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactEstudiantesPos	Estudiantes_Pos_Id	DimEstudiantesPos	Estudiantes_Pos_Id
FactEstudiantesPos	Carrera_Id	DimCarrera	Carrera_Id
FactEstudiantesPos	Facultad_Id	DimFacultad	Facultad_Id
FactEstudiantesPos	Sexo_Id	DimSexo	Sexo_Id
FactEstudiantesPos	Fecha_Inicio_Tiempo_Id	DimTiempo	Tiempo_Id
FactEstudiantesPos	Fecha_Fin_Dim_Tiempo_Id	DimTiempo	Tiempo_Id
FactEstudiantesPos	Fecha_Grado_Dim_Tiempo_Id2	DimTiempo	Tiempo_Id
FactEstudiantesPos	Campus_Id	DimCampus	Campus_Id
FactEstudiantesPos	Periodo_Id	DimPeriodo	Periodo_Id

Anexo 54. Tablas de Hechos Biblioteca

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactBiblioteca	Biblioteca_Id	DimBiblioteca	Proyectos_Vinculacion_Id
FactBiblioteca	Fecha_Reporte_Tiempo_Id	DimTiempo	Tiempo_Id

Anexo 55. Tabla de Hechos Aulas

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactCapacidadAula	Fecha_Reporte_Tiempo_Id	DimTiempo	Tiempo_Id
FactCapacidadAula	Aula_Id	DimAulas	Aula_Id
FactCapacidadAula	Campus_Id	DimCampus	Campus_Id

Anexo 56. Tablas de Hechos Libros Artículos

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactLibroArticulo	Publicacion_Id	DimPublicacion Articulo	Publicacion_Id
FactLibroArticulo	Articulo_Capitulo_Id	DimLibroCapitulo	Articulo_Capitulo_Id
FactLibroArticulo	Base_Datos_Revista_Id	DimBaseDatos RevistaIndexada	Base_Datos_Revista_Id
FactLibroArticulo	Anio_Publicacion_Tiempo_Id	DimTiempo	Tiempo_Id

Anexo 57. Tabla de Hechos Periodo Sabático

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactPeriodoSabatico	Periodo_Sabatico_Id	DimPeriodoSabatico	Periodo_Sabatico_Id
FactPeriodoSabatico	Campus_Id	DimCampus	Campus_Id
FactPeriodoSabatico	Docente_Id	DimDocentes	Docente_Id
FactPeriodoSabatico	Fecha_Reporte_Pac_Tiempo_Id	DimTiempo	Tiempo_Id

Anexo 58. Tabla de Hechos Docentes

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactHorasDocente	Tiempo_Dedicacion_Id	DimTiempoDedicacion	Tiempo_Dedicacion_Id
FactHorasDocente	Ingreso_Concurso_Id	DimIngresoConcurso	Ingreso_Concurso_Id
FactHorasDocente	Subarea_Ultimo_Titulo_Cuarto_Id	DimSubareaUltimoTituloCuarto	Subarea_Ultimo_Titulo_Cuarto_Id
FactHorasDocente	Cargo_Autoridad_Id	DimCargoAutoridad	Cargo_Autoridad_Id
FactHorasDocente	Ultimo_Titulo_Cuarto_Nivel_Id	DimUltimoTituloCuartoNivel	Ultimo_Titulo_Cuarto_Nivel_Id
FactHorasDocente	Categoria_Personal_Academico_Id	DimCategoriaPersonalAcademico	Categoria_Personal_Academico_Id
FactHorasDocente	Ultimo_Titulo_Tercer_Nivel_Id	DimUltimoTituloTercerNivel	Ultimo_Titulo_Tercer_Nivel_Id
FactHorasDocente	Docente_Id	DimDocente	Docente_Id
FactHorasDocente	Campus_Id	DimCampus	Campus_Id
FactHorasDocente	Sexo_Id	DimSexo	Sexo_Id
FactHorasDocente	Fecha_Ingreso_les_Tiempo_Id	DimMetaLogrosAlcanzados	Logros_Alcanzados_Id
FactHorasDocente	Fecha_Fin_les_Tiempo_Id	DimTiempo	Tiempo_Id
FactHorasDocente	Fecha_Inicio_Cargo_Autoridad_Dim_Tiempo_Id2	DimTiempo	Tiempo_Id
FactHorasDocente	Fecha_Fin_Cargo_Autoridad_Dim_Tiempo_Id3	DimTiempo	Tiempo_Id
FactHorasDocente	Fecha_Reporte_Pac_Tiempo_Id	DimTiempo	Tiempo_Id

Anexo 59. Tabla de Hechos Estudiantes Pregrado

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactEstudiantesPre	Estudiante_Id	DimEstudiantesPre	Estudiante_Id
FactEstudiantesPre	Sexo_Id	DimSexo	Sexo_Id
FactEstudiantesPre	Modalidad_Id	DimModalidad	Modalidad_Id
FactEstudiantesPre	Facultad_Id	DimFacultad	Facultad_Id
FactEstudiantesPre	Carrera_Id	DimCarrera	Carrera_Id
FactEstudiantesPre	Fecha_Ulti_Matri_Tiempo_Id	DimTiempo	Tiempo_Id
FactEstudiantesPre	Fecha_Primer_Nivel_Dim_Tiempo_Id	DimTiempo	Tiempo_Id
FactEstudiantesPre	Fecha_Egresada_Dim_Tiempo_Id2	DimTiempo	Tiempo_Id
FactEstudiantesPre	Fecha_Grado_Dim_Tiempo_Id3	DimTiempo	Tiempo_Id
FactEstudiantesPre	Colegio_Id	DimColegio	Colegio_Id
FactEstudiantesPre	Discapacida_Id	DimDiscapacida	Discapacida_Id
FactEstudiantesPre	Localidad_Id	DimLocalidadOrigen	Localidad_Id
FactEstudiantesPre	Campus_Id	DimCampus	Campus_Id
FactEstudiantesPre	Periodo_Id	DimPeriodo	Periodo_Id
FactEstudiantesPre	Estado_Civil_Id	DimEstadoCivil	Estado_Civil_Id

Anexo 60. Tabla de Hechos Proyectos de Investigación

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactProyectosInvestigacion	Proyectos_Investigacion_Id	DimProyectosInvestigacion	Proyectos_Investigacion_Id
FactProyectosInvestigacion	Fecha_Inicio_Proyecto_Tiempo_Id	DimTiempo	Tiempo_Id
FactProyectosInvestigacion	Fecha_Fin_Proyecto_Dim_Tiempo_Id	DimTiempo	Tiempo_Id
FactProyectosInvestigacion	Campus_Id	DimCampus	Campus_Id

Anexo 61. Tabla de Hechos Asignatura Docentes Pregrado

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactAsignaturaDoc	Periodo_Id	DimPeriodo	Periodo_Id
FactAsignaturaDoc	Campus_Id	DimCampus	Campus_Id
FactAsignaturaDoc	Carrera_Id	DimCarrera	Carrera_Id
FactAsignaturaDoc	Materia_Id	DimMateria	Materia_Id
FactAsignaturaDoc	Facultad_Id	DimFacultad	Facultad_Id
FactAsignaturaDoc	Fecha_Inicio_Asig_Tiempo_Id	DimTiempo	Tiempo_Id
FactAsignaturaDoc	Fecha_Fin_Asig_Tiempo_Id	DimTiempo	Tiempo_Id
FactAsignaturaDoc	Docente_Id	DimDocente	Docente_Id

Anexo 62. Tabla de Hechos Cursando Estudios

Tabla destino	Campo destino	Tabla origen	Campo Origen
FactCursandoEstudios	Cursando_Estudios_Id	DimCursandoEstudios	Cursando_Estudios_Id
FactCursandoEstudios	Docente_Id	DimDocentes	Docente_Id
FactCursandoEstudios	Sexo_Id	DimSexo	Sexo_Id
FactCursandoEstudios	Tiempo_Fecha_Pac_Docente_Cursando_Id	DimTiempo	Tiempo_Id