

**DISEÑO DE UN *DATAMART* DE INFORMACIÓN ACADÉMICA PARA LA
UNIVERSIDAD DE NARIÑO**

**JOHN JAIRO BARRETO ARROYO
DIEGO ALEXANDER BOCANEGRA PORTILLA**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2011**

**DISEÑO DE UN *DATAMART* DE INFORMACIÓN ACADÉMICA PARA LA
UNIVERSIDAD DE NARIÑO**

**JOHN JAIRO BARRETO ARROYO
DIEGO ALEXANDER BOCANEGRA PORTILLA**

**Trabajo de Grado presentado como requisito parcial para optar por el título
de Ingeniero de Sistemas**

**Director:
Ricardo Timarán Pereira Ph. D.**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2011**

NOTA DE RESPONSABILIDAD

Las ideas y conclusiones aportadas en el siguiente trabajo son responsabilidad exclusiva del autor.

Artículo 1^{ro} del Acuerdo No. 324 de octubre 11 de 1966 emanado del Honorable Consejo Directivo de la Universidad de Nariño.

Nota de aceptación:

Firma del presidente del jurado

Firma del jurado

Firma del jurado

San Juan de Pasto 5 de Octubre de 2011

DEDICATORIA

A Dios, porque en su infinita sabiduría y misericordia me doto este conocimiento maravilloso, una ilusión y una gran esperanza. Me dio la inspiración para ir en busca de mis sueños y mis metas.

A mi abuela por su apoyo incondicional y por la confianza depositada para el logro de esta meta.

A mi madre por su confianza y su apoyo durante el sendero hacia este logro.

A mi hermana por creer en mí y apoyarme incondicionalmente en el camino a esta meta.

A mis docentes, por su sabiduría, conocimiento y ejemplo, porque me motivaron a continuar.

A todos quienes estuvieron cerca y me brindaron toda la colaboración que necesité.

John Jairo Barreto Arroyo

DEDICATORIA

A Dios , por iluminarme en el desarrollo y finalización de este proyecto.

A mi madre, como reconocimiento al esfuerzo y apoyo que siempre me brindó en el proceso de mi formación como profesional

Diego Alexander Bocanegra Portilla

AGRADECIMIENTOS

Agradecemos a Dios, porque su Espíritu nos llenó de sabiduría, por su gran incondicional presencia en todo momento, por su precioso amor, por darnos la fortaleza, entereza y energía para alcanzar esta meta.

A nuestra familia por su apoyo, comprensión y colaboración incondicional y permanente.

Al ingeniero Ricardo Timarán Pereira Ph. D., Asesor de nuestro proyecto, por su tiempo, colaboración, paciencia y asistencia continua durante el desarrollo de este proyecto.

A la Oficina de Informática de la Universidad de Nariño por su colaboración y atención.

A la vicerrectoría de investigaciones y postgrados VIPRI por el apoyo financiero de este proyecto.

A nuestros amigos y compañeros, por ser partícipes en el proceso de esta etapa de nuestra vida.

A nuestros profesores por compartir con nosotros sus conocimientos y experiencias.

Y a todas aquellas personas que contribuyeron en la realización de este proyecto de investigación.

RESUMEN

EN ESTE PROYECTO DE INVESTIGACIÓN SE PRESENTA EL “**DISEÑO DE UN DATAMART DE INFORMACIÓN ACADÉMICA PARA LA UNIVERSIDAD DE NARIÑO**”. ASÍ COMO EL PROTOTIPO FUNCIONAL DE ESTA INVESTIGACIÓN.

EL DISEÑO ESTA COMPUESTO POR CUATRO ÁREAS ESPECÍFICAS: LA PRIMERA INSCRIPCIONES Y ADMISIONES, LA SEGUNDA MATRÍCULAS Y POBLACIÓN ESTUDIANTIL, LA TERCERA MATERIAS Y RENDIMIENTO ACADÉMICO, Y POR ULTIMO EL ÁREA DE PROFESORES.

EN EL ÁREA DE INCRIPCIONES Y ADMISIONES SE TIENE EN CUENTA INFORMACIÓN DE: INSCRIPCIONES, ADMISIONES Y COBERTURA. EN EL ÁREA DE MATRICULAS Y POBLACIÓN ESTUDIANTIL LAS MATRICULAS Y EL NÚMERO DE ESTIDIANTES, PARA EL AREA MATERIAS Y RENDIMIENTO ACADÉMICO SE TIENEN EN CUENTA: NOTAS DE ALUMNOS E INFORMACIÓN DE MATRICULAS DE MATERIAS, TANTO DE CANCELACIONES, COMO DE MATRÍCULAS. Y PARA ÁREA DE PROFESORES SE ANALIZA LA INFORMACIÓN DE PROFESORES Y EVALUACIÓN DOCENTE.

PARA EL PROTOTIPO FUNCIONAL SE UTILIZA LA HERRAMIENTA *STARCUBE*, LA CUAL PERMITE EL ANÁLISIS MULTIDIMENSIONAL Y LA MANIPULACIÓND E TODA LA INFORMACIÓN ACADÉMICA CONTENIDA EN *ELDATAMART*.

EN ESTE PROYECTO SE PLANTEA DISEÑAR UN *DATAMART* CON LA INFORMACIÓN HISTÓRICA DE LA OFICINA DE REGISTRO ACADÉMICO DE LA UNIVERSIDAD DE NARIÑO, ASÍ COMO EL PROTOTIPO FUNCIONAL DE ESTE. ESTE PROYECTO SE CONSTITUYE COMO EL PRIMER PASO PARA CONTAR EN UN FUTURO CON LA IMPLEMENTACIÓN DE UN *DATAMART* ACADÉMICO QUE SE CONVIERTA EN UN MEDIO EFICAZ DE SOPORTE A LA TOMA DE DECISIONES PARA LOS DIRECTIVOS DE LA UNIVERSIDAD DE NARIÑOEN TODOS LOS ASPECTOS ACADÉMICOS, DE UNA MANERA ACERTADA.

ABSTRACT

IN THIS RESEARCH PROJECT SHOWS THE "**DESIGN OF A DATAMART ACADEMIC INFORMATION FOR THE UNIVERSITY OF NARIÑO**". AND THE WORKING PROTOTYPE OF THIS RESEARCH.

DESIGN CONSISTS OF FOUR SPECIFIC AREAS: THE FIRST REGISTRATION AND ADMISSIONS THE SECOND TUITION AND STUDENT POPULATION THE THIRD SUBJECT AND ACADEMIC PERFORMANCE, AND FINALLY THE AREA OF TEACHERS.

IN THE REGISTRATION AREA AND TAKES INTO ACCOUNT ADMISSIONS INFORMATION: REGISTRATION, ADMISSIONS AND COVERAGE. THE AREA OF TUITION AND STUDENT POPULATION ENROLLMENTS AND THE NUMBER OF STUDENTS. FOR THE ACADEMIC AREA MATTERS ARE TAKEN INTO ACCOUNT: NOTES FROM TUITION STUDENTS AND INFORMATION MATERIALS, BOTH OF CANCELLATIONS, AND ENROLLMENTS. THE TEACHER'S AREA LOOKS INFORMATION FOR TEACHERS AND TEACHER EVALUATION.

TO THE FUNCTIONAL PROTOTYPE STARCUBE TOOL IS USED, WHICH ALLOWS THE ANALYSIS AND MANIPULATION OF MULTIDIMENSIONAL ALL ACADEMIC INFORMATION CONTAINED IN THE DATAMART.

THIS PROJECT PROPOSES TO DESIGN A DATAMART WITH HISTORICAL INFORMATION OF THE ACADEMIC REGISTER OF THE UNIVERSITY OF NARIÑO, AS MUCH AS THE FUNCTIONAL PROTOTYPE ITSELF. THIS PROJECT CONSTITUTES THE FIRST STEP TO HAVING A FUTURE WITH THE IMPLEMENTATION OF AN ACADEMIC DATAMART TO BECOME AN EFFECTIVE MEANS TO SUPPORT DECISION MAKING FOR MANAGERS OF THE UNIVERSITY OF NARIÑO IN ALL ACADEMIC ASPECTS OF A ACCURATELY.

CONTENIDO

	Pág.
INTRODUCCION	23
1. TEMA	25
1.1 TÍTULO	25
1.2 LÍNEA DE INVESTIGACIÓN.....	25
1.3 ALCANCE Y DELIMITACIÓN	25
1.4 PROBLEMA OBJETO DE ESTUDIO	25
1.4.1 Descripción del problema.....	25
1.4.2 Formulación del problema.....	26
1.5 OBJETIVOS	27
1.5.1 Objetivo general.	27
1.5.2 Objetivos Específicos	27
1.6 JUSTIFICACIÓN	27
1.7 ORGANIZACIÓN DEL DOCUMENTO	28
2. DATA WAREHOUSE	29
2.1 CLASIFICACIÓN DE LOS SISTEMAS DE INFORMACIÓN	29
2.1.1 Sistemas Estratégicos.....	29
2.1.2 Sistemas Tácticos	29
2.1.3 Sistemas Operacionales	29
2.2 SISTEMAS DE SOPORTE PARA LA TOMA DE DECISIONES	30
2.2.1 Ventajas y desventajas de un Sistema de Soporte de Decisiones.....	30
2.3 DATAWAREHOUSE Y SUS COMPONENTES	31
2.3.1 Definición.	31
2.3.2 Componentes.....	31
2.3.3 Ventajas y Desventajas.....	33
2.4 DATAMARTS	33

2.4.1	Definición	34
2.4.2	Características.	34
2.4.3	Tipos de Datamart.....	34
2.5	OLTP Y OLAP	36
2.5.1	On-line Transactional processing (OLTP).	36
2.5.2	On-line Analytical Processing (OLAP).....	37
2.6	MODELO MULTIDIMENSIONAL	37
2.6.1	Estructura.....	37
2.6.2	Modelado conceptual de un <i>Data Warehouse</i>	39
2.6.3.2	MOLAP.....	43
2.6.3.4	HOLAP	46
2.6.4	Operaciones sobre el modelo multidimensional.....	47
2.8.2.3	Modelado dimensional	58
2.8.2.4	Procesos de extracción, transformación y carga.....	64
2.8.2.5	Selección e instalación de los productos.....	65
2.8.2.6	Mantenimiento y crecimiento de un <i>Data Warehouse</i>	66
3.	DISEÑO DEL DATAMART DE INFORMACIÓN ACADÉMICA	67
3.1	PLANEACIÓN Y ADMINISTRACIÓN DEL PROYECTO.....	67
3.1.1	Objetivo del proyecto.....	67
3.1.2	Definición del proyecto	67
3.1.3	Alcance del Proyecto.....	67
3.1.4	Justificación del proyecto	68
3.2	ANÁLISIS DE REQUERIMIENTOS	68
3.2.1	Documentación de requerimientos.....	68
3.2.1.1	Área de inscripciones y admisiones	68
3.2.1.2	Área de matrículas y población estudiantil.....	70
3.2.1.3	Área de materias y rendimiento académico	73
3.2.1.4	Área de profesores.....	75
3.3	MODELADO DIMENSIONAL	76
3.3.1	El Datamart.	77

3.3.2	Definición de granularidad.....	81
3.3.2.1	Granularidad área de inscripciones y admisiones	81
3.3.2.2	Granularidad área de matrículas y población estudiantil.....	81
3.3.2.3	Granularidad área de matrículas de materias y rendimiento académico.....	81
3.3.3	Dimensiones	83
3.3.3.1	Dimensiones comunes en el Datamart.....	83
3.3.3.2	Dimensiones área de inscripciones y admisiones.....	87
3.3.3.3	Dimensiones de matrículas y población estudiantil.....	88
3.3.3.4	Dimensiones de materias y rendimiento académico.....	89
3.3.3.5	Dimensiones área de profesores	90
3.3.4	Tablas de hechos.....	92
3.3.4.1	Hechos inscripciones y admisiones.....	92
3.3.4.2	Hechos matrículas y población estudiantil	92
3.3.4.3	Hechos matrículas de materias y rendimiento académico	94
3.3.4.4	Hechos profesores	97
3.3.5	Diseño del Modelo Dimensional.....	98
3.3.5.1	Modelo dimensional inscripciones y admisiones.....	98
3.3.5.2	Modelado dimensional matrículas y población estudiantil.....	98
3.3.5.3	Modelado dimensional materias y rendimiento académico.....	98
3.3.5.4	Modelado dimensional profesores.....	101
3.4	DISEÑO TÉCNICO DE LA ARQUITECTURA.....	103
3.4.1	Datos.....	103
3.4.1.1	Mapeo dimensiones comunes.....	105
3.4.1.2	Mapeo inscripciones y admisiones.....	113
3.4.1.3	Mapeo matrículas y población estudiantil.....	118
3.4.1.4	Mapeo de materias y rendimiento	120
3.4.1.5	Mapeo profesores.....	127
3.4.2	Back Room.....	132
4.	PRUEBAS Y RESULTADOS	134

4.1	PROTOTIPO DEL <i>DATAMART</i> DE INFORMACIÓN ACADÉMICA....	134
4.2	HERRAMIENTA DE ANÁLISIS MULTIDIMENSIONAL STARCUBE ..	135
4.2.1	Instalación STARCUBE.....	136
4.3	CUBO DE DATOS DEL ÁREA DE INSCRIPCIONES Y ADMISIONES.....	137
4.4	MANEJO DE LA HERRAMIENTA STARCUBE SOBRE EL ÁREA DE INSCRIPCIONES Y ADMISIONES	143
4.5.4	Información de profesores.....	159
5.	CONCLUSIONES	163
6.	RECOMENDACIONES	165
7.	BIBLIOGRAFIA	166

LISTA DE TABLAS

	Pág.
Tabla 1. Ventajas y desventajas de un <i>DDS</i>	30
Tabla 2. Ventajas y desventajas de un <i>Data Warehouse</i>	33
Tabla 3: Bases de la arquitectura de <i>Data Warehouse</i>	61
Tabla 4. Requerimientos inscripciones y admisiones.....	68
Tabla 5. Requerimientos matrículas e información de estudiantes.....	71
Tabla 6. Requerimientos de matrículas de materias.....	73
Tabla 7. Requerimientos de rendimiento académico.....	74
Tabla 8. Requerimientos de profesores.....	75
Tabla 9. Dimensión Tiempo.....	83
Tabla 10. Dimensión alumno.....	84
Tabla 11. Dimensión carrera.....	85
Tabla 12. Dimensión colegio.....	85
Tabla 13. Dimensión región.....	86
Tabla 14. Dimensión materias.....	86
Tabla 15. Dimensión estado.....	86
Tabla 16. Dimensión Inscrito.....	87
Tabla 17. Dimensión admitido.....	88
Tabla 18. Dimensión semestre.....	88
Tabla 19. Dimensión vigente.....	89
Tabla 20. Dimensión aprobación.....	89
Tabla 21. Dimensión eventos.....	90
Tabla 22. Dimensión eventos materias.....	90
Tabla 23. Dimensión profesor.....	91
Tabla 24. Dimensión nivel académico.....	91
Tabla 25. Dimensión vinculación.....	91

Tabla 26. Dimensión pregunta	92
Tabla 27. Tabla de hechos cobertura.....	93
Tabla 28. Tabla de hechos matrículas y población estudiantil	93
Tabla 29. Tabla de hechos rendimiento académico y materias matriculadas	94
Tabla 30. Tabla de hechos eventos sobre materias.....	95
Tabla 31. Tabla de hechos eventos sobre estudiantes	96
Tabla 32. Tabla de hechos cancelaciones	96
Tabla 33. Tabla de hechos profesores	97
Tabla 35. Mapeo Dimensión tiempo.....	106
Tabla 36. Mapeo dimensión alumno	106
Tabla 37. Mapeo dimensión carrera.....	109
Tabla 38. Dimensión colegio	110
Tabla 39. Mapeo dimensión región	111
Tabla 40. Mapeo dimensión materias	111
Tabla 41. Mapeo dimensión estado	112
Tabla 42. Operaciones de la tabla de hechos cobertura	113
Tabla 43. Mapeo Dimensión Inscrito	115
Tabla 44. Mapeo dimensión admitido.....	118
Tabla 45. Operaciones de la tabla de hechos matrículas y población estudiantil	119
Tabla 46. Mapeo dimensión vigente.....	119
Tabla 47. Mapeo dimensión semestre	120
Tabla 48. Operaciones de la tabla de hechos rendimiento académico y materias matriculadas.....	121
Tabla 49. Operaciones de la tabla de hechos eventos sobre materias	122
Tabla 50. Operaciones de la tabla de hechos de eventos sobre estudiantes.	123
Tabla 51. Operaciones de la tabla de hechos cancelaciones	124
Tabla 52. Mapeo Dimensión aprobación.....	124
Tabla 53. Mapeo dimensión estado	125
Tabla 54. Mapeo dimensión eventos.....	126
Tabla 55. Mapeo dimensión eventos materias	126

Tabla 56. Operaciones de la tabla de hechos profesores	127
Tabla 57. Operaciones de la tabla de hechos evaluaciones	128
Tabla 58. Mapeo dimensión profesor	130
Tabla 59. Mapeo dimensión nivel académico	131
Tabla 60. Mapeo dimensión vinculación	131
Tabla 61. Mapeo dimensión pregunta	132
Tabla 62. Información de inscripciones detallada por año	149
Tabla 63. Información de inscripciones y admisiones del programa de psicología	150
Tabla 64. Información de inscripciones y admisiones por programas	151
Tabla 65. Información de inscripciones y admisiones de Pasto e Ipiales	152
Tabla 66. Población estudiantil por año	152
Tabla 67. Información de estado de estudiantes de Ingeniería de Sistemas.	153
Tabla 68. Información matriculas ingeniería de sistemas primer semestre	154
Tabla 69. Notas y materias matriculadas por año	155
Tabla 70. Información de eventos sobre estudiantes.	156
Tabla 71. Información de eventos sobre materias.....	158
Tabla 72. Información de cancelaciones 2010/2011	159
Tabla 73. Información de profesores por departamento.....	159
Tabla 74. Información de vinculaciones de profesores.	160
Tabla 75. Información de evaluaciones.....	161

LISTA DE FIGURAS

	Pág.
Figura 1. Clasificación de los sistemas de información	29
Figura 2. Componentes de un <i>Data Warehouse</i>	32
Figura 3. <i>Data Warehouse</i> compuesto por varios <i>Datamarts</i>	34
Figura 4. <i>Datamarts</i> dependientes	35
Figura 5. <i>Datamarts</i> independientes	36
Figura 6. Estructura del modelo multidimensional.....	38
Figura 7. Esquema estrella	40
Figura 8. Esquema de constelación	41
Figura 9. Esquema de copo de nieve.....	42
Figura 10. Cubo de datos Multidimensional.	43
Figura 11. Niveles ROLAP	45
Figura 12. Drill Down.....	47
Figura 13. <i>Roll up</i>	48
Figura 14. Slice & dice	48
Figura 15. <i>Pivot</i>	49
Figura 16. Sistema ETL.....	50
Figura 17. Proceso del modelo Inmon.....	53
Figura 18. Tareas de la metodología Kimball.....	55
Figura 19. Proceso de Modelado Dimensional (Kimball).....	59
Figura 20. Cubo inscripciones y admisiones	77
Figura 21. Cubo matrículas y población estudiantil	78
Figura 22. Cubo Materias y Rendimiento académico.....	78
Figura 23. Cubo cancelaciones.....	79
Figura 24. Cubo eventos estudiantes.....	79
Figura 25. Cubo eventos materias	80

Figura 26. Cubo Profesores	80
Figura 27. Cubo Evaluaciones	81
Figura 28. Cubo inscripciones y admisiones detallado.....	99
Figura 29. Cubo de matrículas y población estudiantil detallado	99
Figura 30. Cubo materias y rendimiento académico detallado.....	100
Figura 31. Cubo materias canceladas detallado	100
Figura 32. Cubo eventos estudiantes detallado	101
Figura 33. Cubo eventos materias detallado.....	101
Figura 34. Cubo de profesores detallado	102
Figura 36. Modelo E-R área de inscripciones y admisiones.....	103
Figura 37. Modelo E-R área de matrículas y población estudiantil.....	104
Figura 38. Modelo E-R área de materias y rendimiento académico.....	104
Figura 39. Modelo E-R área de profesores.	105
Figura 40. Inicio de sesión STARCUBE	136
Figura 41. Pantalla inicial STARCUBE	137
Figura 42. Creación de un cubo de datos	138
Figura 43. Añadir / Editar una conexión.	138
Figura 44. Creación de la estructura cubo de datos inscripciones y admisiones.	139
Figura 45. Selección de medidas	140
Figura 46. Agregar dimensiones y jerarquías.....	141
Figura 47. Esquema XML del cubo de datos.	142
Figura 48. Almacenamiento de un Cubo.....	143
Figura 49. Cubos creados STARCUBE.....	144
Figura 50. Navegación sobre un cubo OLAP.	145
Figura 51. Tabla OLAP.....	145
Figura 52. Diagrama tipo pastel inscripciones y admisiones.....	146
Figura 53. Restricción medidas o métricas.	146
Figura 54. Operación <i>dice</i>	147
Figura 55. Operación <i>pivote</i>	148
Figura 56. Operación <i>drilldown</i>	148

LISTA DE ANEXOS

	Pág.
ANEXO A	169
ANEXO B	170
ANEXO C.....	172

GLOSARIO

Ad hoc. Es una herramienta elaborada específicamente para una determinada ocasión o situación, permite al usuario personalizar una consulta en tiempo real.

Back Room. Área responsable de extraer y preparar los datos en el *Data Warehouse*.

Cubo de datos. Formato multidimensional en el que las celdas contienen datos numéricos, llamados medidas, organizados por temas, denominados dimensiones.

Datamart. *Data Warehouse* a menor escala. Está orientado a un área específica del negocio. Todas sus métricas y dimensiones están relacionadas con un área de negocio en particular.

Data Warehouse. Colección de datos orientados al tema, integrados y no volátiles para la toma de decisiones.

DBMS. Sistema de gestión de bases de datos. Software que controla la organización, almacenamiento, recuperación, seguridad e integridad de los datos en una base de datos.

Dice. Operación OLAP de restricción de los datos del cubo.

Dimensión. Categorías descriptivas de un cubo de datos. Se encarga de agrupar, calificar o catalogar cada uno de los hechos contenidos dentro de un Datamart.

Drill Down. Operación OLAP para aumentar el nivel de detalle de los datos del cubo.

DSS. Sistemas de soporte a las decisiones, permiten extraer la información estratégica de la compañía y, mediante la aplicación de técnicas de análisis de éstas, saber qué resultados generarán unas u otras decisiones.

ETL. Proceso de extracción, transformación y carga de datos desde una fuente determinada hacia un *Datamart* o el *Data Warehouse* corporativo.

Front End. Hace referencia al estado inicial de un proceso. Es el sistema de software que interactúa directamente con el usuario responsable de recoger entradas de los usuarios, y ser procesadas.

Front Room. Es el área del Ambiente de Usuario Final, Las Herramientas de Consultas.

Granularidad. Describe el nivel de detalle de la base de datos en un *Data Warehouse*. La granularidad es el factor que determina el nivel de *Drill Down* podemos navegar.

HOLAP. Procesamiento analítico híbrido en línea. Clasificación de OLAP en donde se hace uso de DBMS relacionales y DBMS multidimensionales.

Inteligencia de negocios. Es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

Jerarquía. Es como se organizan las dimensiones. Una dimensión puede contener distintas jerarquías.

Métrica. Valores cuantitativos de un cubo, a los que se pueden operar aritméticamente.

MOLAP. Procesamiento analítico multidimensional en línea, tecnología diseñada para realizar análisis de datos a través de un modelo de datos multidimensional.

OLAP. Sistemas de procesamiento analítico en línea, define a una tecnología que se basa en el análisis multidimensional de los datos y que le permite al usuario tener una visión más rápida e interactiva de los mismos.

OLTP. Sistemas de procesamiento de transacciones en línea, o sistemas transaccionales, en los cuales residen las operaciones del día a día de cada negocio y que son la fuente prioritaria de datos para cada *Datamart* o el *Data Warehouse* corporativo.

RDBMS. Sistemas de gestión de bases de datos relacionales. Constituyen la base para los sistemas OLTP.

ROLAP. Procesamiento analítico relacional en línea. Clasificación de OLAP para DBMS relacionales.

Roll Up. Operación OLAP para disminuir el nivel del detalle de los datos del cubo.

Sistemas legacy. Sistemas fuente de la empresa con los que manejan la operación de las transacciones diarias. Su característica principal es el enfoque en los procesos de operación diaria de la empresa.

Slice. Operación OLAP para agregar o quitar una dimensión en el análisis del cubo.

Tabla de hechos.Tabla de datos central que contiene cada una de las métricas que se desea medir dentro de un área específica del negocio.

Trigger.Disparador. Se define así a una subrutina que es ejecutada de manera automática cuando se produce algún tipo de transacción (inserción, borrado o actualización) en la tabla de una base de datos.

INTRODUCCIÓN

El objetivo de las empresas es aprovechar la información que se maneja dentro de ellas, generando conocimiento, con el fin de tener éxito en el área donde se desempeñan. Sin embargo, la información en bruto, no ayuda a mejorar el conocimiento sobre la situación actual de la organización, convirtiéndose en un elemento muchas veces distractor y que no agrega valor. La inteligencia de negocios permite obtener conocimiento a través de información importante de una organización, para la toma de decisiones y desarrollo de estrategias a partir de los datos existentes. Igualmente, el proceso de inteligencia de negocios permite aprovechar el conocimiento obtenido, para un mejor ejercicio de las áreas administrativas, financieras, técnicas y operativas dentro de las instituciones, dando un mejor soporte a la toma de decisiones.

La información que se trata en las organizaciones se obtiene a través de datos almacenados en diversas fuentes, entre ellas las bases de datos transaccionales, las cuales son vitales en el funcionamiento de la empresa.

Para responder a las necesidades de información se requiere de datos históricos y datos de otras fuentes externas e internas. Existen estructuras de datos denominadas *Data Warehouse* o Bodegas de Datos, que permiten trabajar con grandes volúmenes de datos que varían con el tiempo, en las cuales se puede unificar la información de la organización, de manera histórica y organizada, con el fin de extraer conocimiento útil para la toma de decisiones, permitiendo hacer consultas y análisis de la información de la empresa de una manera rápida y sencilla. El *Data Warehouse* integra así mismo la información pertinente a todas las áreas de una empresa, en estructuras internas de datos denominadas *DataMarts*. Dicho de otra manera un *Data Warehouse* está compuesto de uno o varios Mercados de datos o *Datamarts*, los cuales soportan la toma de decisiones de cada área específica dentro de la empresa u organización.

Uno de los sistemas utilizados dentro de la Inteligencia de negocios son las herramientas OLAP, con las cuales se realiza un análisis en tiempo real sobre un *Data Warehouse*. Las herramientas OLAP permiten al usuario hacer consultas obteniendo informes y resúmenes de forma rápida, consistente e interactiva a través de diferentes puntos de vista, dando soporte a la toma de decisiones. Mediante el procesamiento OLAP y sus herramientas, un usuario podría analizar por ejemplo resultados de ventas e ir profundizando en detalle en cada región,

canal de venta, equipos de ventas o vendedores, artículos, hasta encontrar el conocimiento importante para la toma de decisiones.

Dentro área académica de la Universidad de Nariño se manejan varias bases de datos transaccionales que contienen grandes volúmenes de información, siendo así muy difícil el soporte a la toma de decisiones de una manera rápida y concreta, ya que no se cuenta con sistemas para soportar la toma de decisiones, obteniendo muchas veces datos innecesarios y obteniendo información que lleva a un análisis ineficiente.

Para dar un mejor soporte a la toma de decisiones en este proyecto se plantea el diseño de una solución de inteligencia de negocios, un *Datamart*, el cual integra la información de las bases de datos de la Oficina de Registro Académico y el Centro de informática, en donde se ha identificado las necesidades y requerimientos de información académica de la Universidad de Nariño, teniendo en cuenta la metodología apropiada para el diseño de *Data Warehouse*.

Al abordar este documento, se expone el conocimiento teórico del proceso de *Data Warehouse*, teniendo en cuenta los conceptos de sistemas de soporte para la toma de decisiones, y los conceptos propios del *Data Warehouse*. Así mismo se tiene en cuenta las metodologías existentes, destacando principalmente la metodología establecida por Ralph Kimball, quien es considerado la autoridad en el mundo del *Data Warehouse*. Además, se describe en forma detallada el proceso de diseño del *Datamart* de información académica, utilizando la metodología mencionada anteriormente. Posteriormente se exponen y analizan las pruebas y resultados sobre un prototipo funcional conforme al diseño planteado, bajo el Sistema Gestor de Bases de Datos *PostgreSQL* y la herramienta de análisis multidimensional *StarCube*, desarrollada en los laboratorios KDD del grupo de investigación GRIAS (GRUPO DE INVESTIGACIÓN APLICADA A SISTEMAS), del Departamento de Sistemas de la Universidad de Nariño. Finalmente se exponen las conclusiones y recomendaciones que ha dejado el proyecto, para dar claridad a los conceptos desarrollados y servir de base a una futura implementación del *Datamart* de otros proyectos similares de *Data Warehouse*.

En este documento se presenta el trabajo de grado para optar por el título de Ingeniero de Sistemas. El resultado final será “DISEÑO DE UN DATAMART DE INFORMACIÓN ACADÉMICA PARA LA UNIVERSIDAD DE NARIÑO”.

1. TEMA

1.1 TÍTULO

DISEÑO DE UN *DATAMART* DE INFORMACIÓN ACADÉMICA PARA LA UNIVERSIDAD DE NARIÑO.

1.2 LÍNEA DE INVESTIGACIÓN

El presente trabajo de grado, se encuentra inscrito bajo la línea de sistemas de soporte a la toma de decisiones del grupo de investigación GRIAS del departamento de sistemas.

1.3 ALCANCE Y DELIMITACIÓN

En este proyecto se plantea diseñar un *Datamart* con la información histórica de la oficina de registro académico de la Universidad de Nariño, así como el prototipo funcional de este. Este proyecto se constituye como el primer paso para contar, en un futuro con la implementación de un *Datamart* académico que se convierta en un medio eficaz que permita a los directivos de la Universidad de Nariño, soportar la toma de decisiones relacionadas con todos los aspectos académico, de una manera acertada.

La metodología de diseño del *Datamart* aplicada en este proyecto es la utilizada por Ralph Kimball para el diseño e implementación de *Data Warehouse*, utilizando una extensión del esquema en estrella, el cual es el esquema en constelación, tomando como fuente la información académica de pregrado dentro de la Universidad de Nariño.

1.4 PROBLEMA OBJETO DE ESTUDIO

1.4.1 Descripción del problema. Muchos investigadores y empresas han observado la necesidad de representar la información almacenada en bases de datos de una manera fácil de interpretar, para la toma de decisiones de nivel estratégico, haciendo de esta un área activa de investigación [14].

En la actualidad se manejan Motores de Bases de Datos potentes que soportan Bases de Datos robustas y dinámicas, permitiendo realizar transacciones a través

de redes locales o externas, sistemas que hacen parte de las Intranet, Extranet e Internet. Estos procesos complejos generan gran cantidad de datos transaccionales y por ende crecimiento de los archivos de las Bases de Datos, pero la mayoría de procesos utilizan esos datos para la información diaria que soportan las gestiones financieras y administrativas de las entidades.

Para el análisis de datos es necesario recurrir a resúmenes que den soporte a aplicaciones, sin utilizar los datos originales, de los cuales se pueden hacer estudios más detallados, en donde entra a actuar el concepto de *Datamart*, cuyo trabajo de análisis es denominado minería de datos.

Las organizaciones, entidades y empresas que manejan el flujo de información utilizando Bases de Datos para agilizar sus procesos, pueden utilizar los datos almacenados para realizar estudios retrospectivos y proyectar acciones futuras aplicando sistemas de soporte de decisión (DSS *DecisionSupportSystem*), para ello es necesario construir el *Data Warehouse* mediante la implementación de *Datamarts*.

Los *Datamarts* son subconjuntos de un *Data Warehouse* y a su vez un conjunto de hechos y datos organizados para el soporte de toma de decisiones basados en la necesidad de un área o departamento específico.

Las bases de datos de la Oficina de Registro y Control académico maneja un gran volumen de datos históricos, pero debido a que no se cuentan con medios ni herramientas de Inteligencia de Negocios, no es posible realizar análisis de datos y estudios retrospectivos de una manera eficaz, que permitan apoyar las decisiones que se toman en esta área.

El diseño del *Datamart* académico será la primera aproximación de esta oficina y de la Universidad de Nariño de contar con sistemas de Inteligencia de Negocios que soporten la toma de decisiones relacionadas con todos los aspectos académicos, de una manera acertada

1.4.2 Formulación del problema. ¿Cómo utilizar el volumen de datos de las bases de datos de la Oficina de Registro de Control Académico de la Universidad de Nariño para soportar la toma de decisiones acertadas de sus directivos?

1.5 OBJETIVOS

1.5.1 Objetivo general. Diseñar una solución de Inteligencia de Negocios para soportar la toma de decisiones de los directivos de la Universidad de Nariño de una manera acertada, mediante un *Datamart* que integre la información de las bases de datos de la Oficina de Registro Académico y el Centro de Informática.

1.5.2 Objetivos Específicos

- Recopilar y Apropiar la información relacionada con *Datamarts*.
- Identificar las necesidades y los requerimientos de información académica de la Universidad de Nariño.
- Evaluar la metodología de diseño de *Data Warehouse* que más se adecue a las necesidades del diseño del *Datamart* académico de la Universidad de Nariño.
- Diseñar el *Datamart* académico de tipo constelación.
- Construir un prototipo y realizar pruebas de funcionalidad del prototipo del *Datamart* académico mediante una herramienta de análisis multidimensional.

1.6 JUSTIFICACIÓN

El presente proyecto permitió a los investigadores obtener una base teórica referente al concepto de *Datamart* y su uso, teniendo en cuenta que el desarrollo de este involucra las características de un *Data Warehouse*, utilizando metodologías para el desarrollo y explotación de *Data Warehouse* como una herramienta para la toma de decisiones.

Permitirá a la Universidad de Nariño dejar un precedente de los resultados de la investigación que pueden ser utilizados por la universidad para financiar la implementación del *Datamart* y aplicar herramientas de minería de datos.

Los resultados están contenidos en el presente documento con toda la información y conocimiento recolectado en el transcurso de la investigación, el documento

quedará dispuesto para el público en general como guía para futuras investigaciones, permitiendo que los mayores beneficiados por la investigación sean los estudiantes de la Universidad de Nariño, así como también los docentes e investigadores que quieran consultar, profundizar o tomar como referencia este proyecto de gran importancia.

1.7 ORGANIZACIÓN DEL DOCUMENTO

Este documento se encuentra organizado de la siguiente manera: en el capítulo 1 se especifican los objetivos generales y específicos del proyecto, en el capítulo 2 se presenta la información preliminar de Data Warehouse y metodologías para un proyecto de Data Warehouse, en el capítulo 3 se muestra el proceso del análisis y diseño del Datamart académico para la Universidad de Nariño utilizando la metodología de Ralph Kimball, en capítulo 4 se presenta las pruebas y resultados sobre el prototipo construido con la herramienta de análisis multidimensional *STARCUBE*, desarrollada en el laboratorio KDD del grupo de investigación GRIAS del programa de Ingeniería de Sistemas de la Universidad de Nariño, en el capítulo 5 las conclusiones y en el capítulo 6 las recomendaciones.

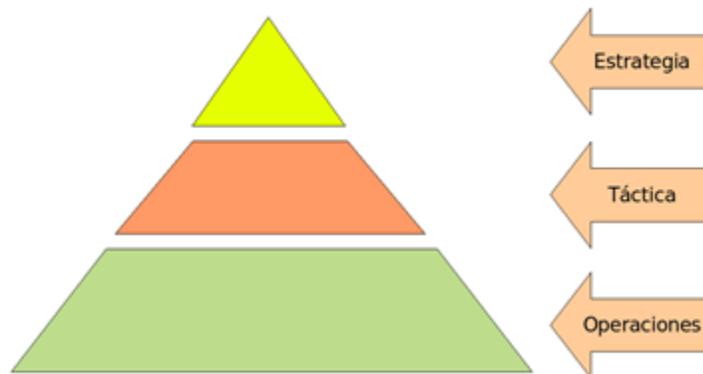
2. DATA WAREHOUSE

En este capítulo se realiza una descripción acerca del proceso de toma de decisiones, *Data Warehouse* y *Datamarts*, además de las metodologías de diseño e implementación de *Data Warehouse*.

2.1 CLASIFICACIÓN DE LOS SISTEMAS DE INFORMACIÓN

Los sistemas de información están divididos de acuerdo al siguiente esquema [12]:

Figura 1. Clasificación de los sistemas de información



Fuente. Timarán Pereira, Ricardo. Bodega de datos y procesamiento analítico en línea. 2009. Pág. 5. [12]

2.1.1 Sistemas Estratégicos. Están orientados a soportar la toma de decisiones, facilitan la dirección, proporcionándole un soporte básico para la toma de decisiones. Se caracterizan porque son sistemas sin carga periódica de trabajo, es decir, su utilización no es predecible.

2.1.2 Sistemas Tácticos. Son diseñados para soportar las actividades de coordinación de actividades y manejo de documentación, definidos para facilitar consultas sobre información almacenada en el sistema, proporcionar informes, en resumen, facilitar la gestión independiente de la información por parte de los niveles intermedios de la organización.

2.1.3 Sistemas Operacionales. Cubren el núcleo de operaciones tradicionales de captura masiva de datos y servicios básicos de tratamiento de datos, con tareas

predefinidas (contabilidad, facturación, almacén, presupuesto, personal y otros sistemas administrativos)[12].

2.2 SISTEMAS DE SOPORTE PARA LA TOMA DE DECISIONES

Existen funciones dentro de la empresa que tienen que ver con el planteamiento, previsión y administración de la organización. Estas funciones son críticas para la supervivencia de la organización, especialmente en un mundo de cambios rápidos.

Las funciones como “planeamiento de ingeniería” o “análisis financiero” son diferentes a las funciones de los sistemas de información operacionales, junto con los tipos de sistemas y sus requerimientos [1]. Los DSS (Sistemas de soporte a la toma de decisiones, *Decision Support Systems*), dan soporte a las personas que toman las decisiones en una organización ofreciéndoles datos de más alto nivel para ayudarles en las decisiones importantes y complejas[3].

Mientras las necesidades de los datos operacionales se enfocan normalmente hacia una sola área, los datos para el soporte a la toma de decisiones, toman un número de áreas diferentes y necesitan cantidades grandes de operaciones relacionadas, para obtener conocimiento, siendo el soporte de decisiones importantes sobre como operará la empresa en el futuro.

La tecnología de *Data Warehouse* basa sus conceptos y diferencias entre dos tipos fundamentales de sistemas de información en todas las organizaciones: los sistemas operacionales y los sistemas de soporte de decisiones. Este último sistema de información es la base de un *Data Warehouse*[1].

2.2.1 Ventajas y desventajas de un Sistema de Soporte de Decisiones: Las ventajas y desventajas de estos sistemas se describen en la siguiente tabla:

Tabla 1. Ventajas y desventajas de un *DDS*.

Ventajas	Desventajas
<ul style="list-style-type: none">• Motivan la exploración y el descubrimiento del tomador de decisiones.	<ul style="list-style-type: none">• Tiene como limitantes: el equipo computacional en el que corre, el diseño del mismo, y el conocimiento que posee cuando

Tabla 1. Continuación.

Ventajas	Desventajas
<ul style="list-style-type: none"> • Extienden la capacidad del tomador de decisiones para procesar la información y el conocimiento. • Permiten reducir el tiempo necesario para llegar a una decisión. • Aunque es difícil evaluar, normalmente incrementan la confiabilidad de los resultados que se logran al tomar una decisión. • Permiten al atacar problemas complejos, de gran escala y que normalmente demandarían muchísimo tiempo sin tener el uso de DSS. • Dan nuevas Evidencias para apoyar una decisión o apoyan las suposiciones previas. • Crean una nueva ventaja competitiva. 	<p>es usado.</p> <ul style="list-style-type: none"> • Las interfaces no son suficientemente sofisticadas como para soportar la generación de voz, el reconocimiento de voz y el manejo del lenguaje natural. • Normalmente no involucran aspectos de inteligencia humana como la creatividad, la imaginación y la intuición. • Son aplicables a contextos específicos.

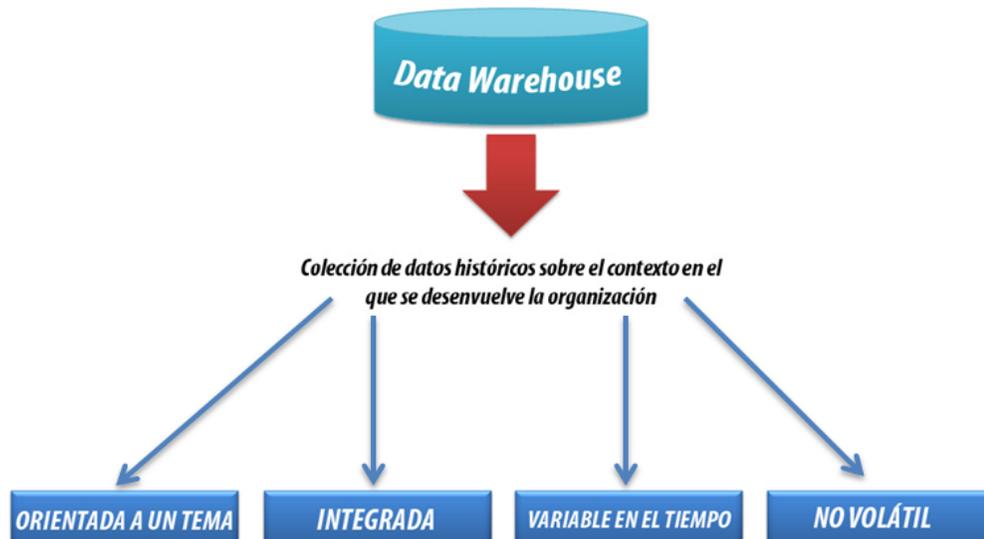
Fuente. Timarán Pereira, Ricardo. Bodega de datos y procesamiento analítico en línea. 2009. Pág. 28. [12].

2.3 DATAWAREHOUSE Y SUS COMPONENTES

2.3.1 Definición. Un *Data Warehouse* es una colección de datos orientados a un tema o sujetos, integrados, variable en el tiempo y no volátiles[6], con el fin de ayudar en la toma de decisiones estratégicas, permitiendo aplicar eficientemente herramientas para resumir, describir y analizar los datos[12].

2.3.2 Componentes. En la figura 2 se observan los diferentes componentes de un *Data Warehouse*.

Figura 2. Componentes de un *Data Warehouse*



Fuente.Timarán Pereira, Ricardo. Bodega de datos y procesamiento analítico en línea. 2009. Pág. 17. [12]

- **Orientados a temas o sujetos.** Se centra en entidades de alto nivel, es decir, en los temas de mayor importancia para la organización como clientes, ventas, productos Integrado[12]. Esta orientación difiere con la orientación de los procesos en el procesamiento de las transacciones [7].
- **Integrados.** Integra datos recogidos de diferentes sistemas operacionales de la organización y/o fuentes externas, proporcionando una base de datos individual y unificada para el apoyo a la toma de decisiones [7], eliminando inconsistencias existentes entre los diversos sistemas operacionales. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios [4].
- **Variable en el tiempo.** Los datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente. En la dimensión tiempo se puede identificar tendencias, pronosticar operaciones futuras y establecer objetivos de operación. Una *Data Warehouse* puede mantener información de más de un instante, almacenando los datos en una extensa serie de fotografías instantáneas, donde cada una representa los datos operativos capturados en un momento de tiempo dado.

- **No volátiles.** La información en el *Data Warehouse* existe para ser leída y no modificada, los datos son históricos y el anexo de datos, sin ningún tipo de acción sobre lo que ya exista, se conoce como refrescar el *Data Warehouse*. Esto está en contraste con la información del sistema transaccional que está sujeta a permanentes inserciones, actualizaciones, reemplazos o borrados.

2.3.3 Ventajas y Desventajas. En la tabla 2 se describe las ventajas y desventajas de un *Data Warehouse*.

Tabla 2. Ventajas y desventajas de un *Data Warehouse*.

Ventajas	Desventajas
<ul style="list-style-type: none"> - Alto retorno de inversión. - Aumento de la competitividad en el mercado. - Aumento de la productividad de los técnicos de dirección. 	<ul style="list-style-type: none"> - Altos costos de mantenimiento. - Subestimación del esfuerzo necesario para su diseño y creación. - Subestimación de los recursos necesarios para la captura, carga y almacenamiento de los datos. - Incremento continuo de los requerimientos de los usuarios - Privacidad de los datos. - Alta demanda de recursos.

Fuente: Timarán Pereira, Ricardo. Bodega de datos y procesamiento analítico en línea. 2009. Pág. 27. [12]

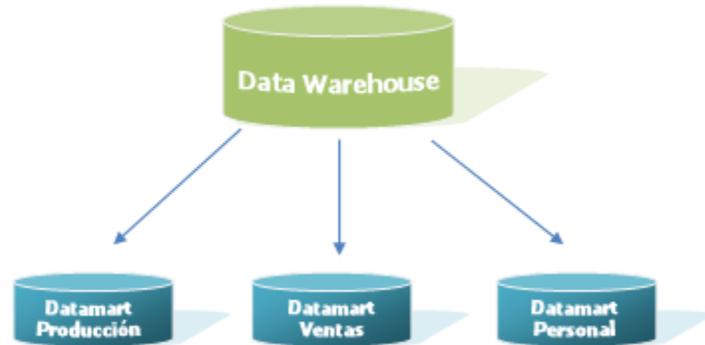
A pesar de los beneficios potenciales de un *Data Warehouse*, muchos proyectos de *Data Warehouse* son grandes esfuerzos que implican coordinación entre muchas partes de una organización. Muchas organizaciones han subestimado el tiempo y el esfuerzo para conciliar diferentes vistas de un *Data Warehouse* puede llevarlo a un desempeño deficiente. Una arquitectura apropiada puede ayudar a aliviar problemas con el desempeño de *Data Warehouse* y el esfuerzo de desarrollo[7].

2.4 DATAMARTS

Es importante organizar la información en diferentes áreas, como por ejemplo departamento de ventas. Siendo más fácil representar la información de esta manera. En un *Data Warehouse* no es posible satisfacer todas las necesidades de

las áreas de una organización en un solo esquema, por esta razón un *Data Warehouse* puede estar compuesto de varios *Datamarts*[5].

Figura 3. *Data Warehouse* compuesto por varios *Datamarts*.



Fuente. Hernández, José y Ramírez, María. Introducción a la Minería de Datos. 2004. Pág. 48.[5]

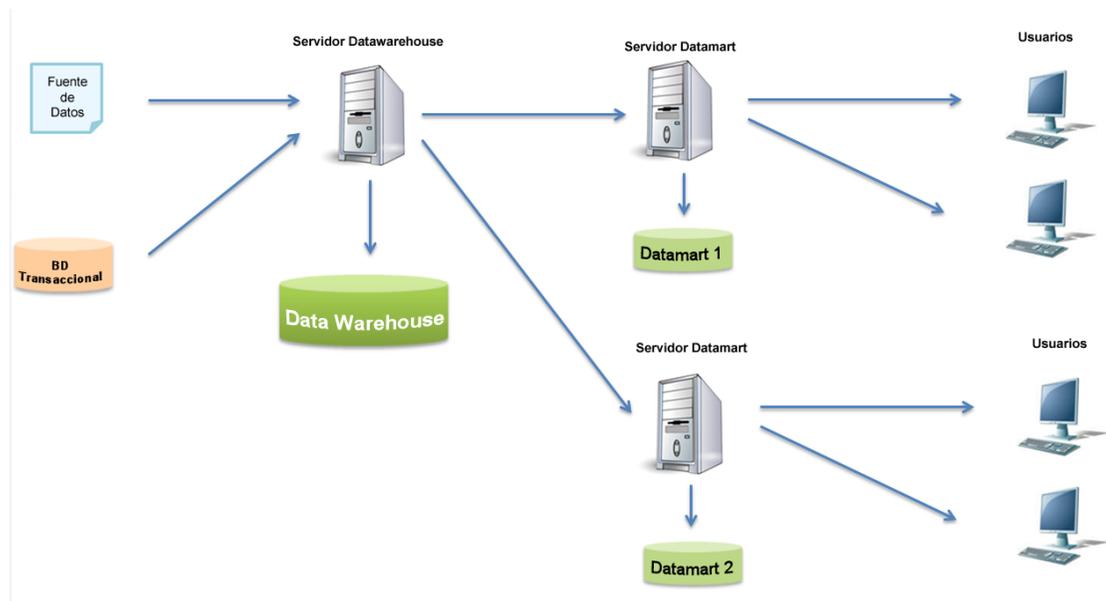
2.4.1 Definición. Un *Datamartes* un subconjunto y/o una especialización de un *Data Warehouse*, enfocado a un departamento o un área específica, como por ejemplo el área de ventas o de contabilidad de una empresa.

2.4.2 Características. Como el *Data Warehouse* tiene las mismas características de integración, no volatilidad y orientación temática.

2.4.3 Tipos de Datamart .Existen dos tipos: los dependientes y los independientes cuya definición se describe a continuación.

Datamarts Dependientes: Se construyen a través de un *Data Warehouse* central, esto indica que reciben sus datos de un repositorio empresarial central.

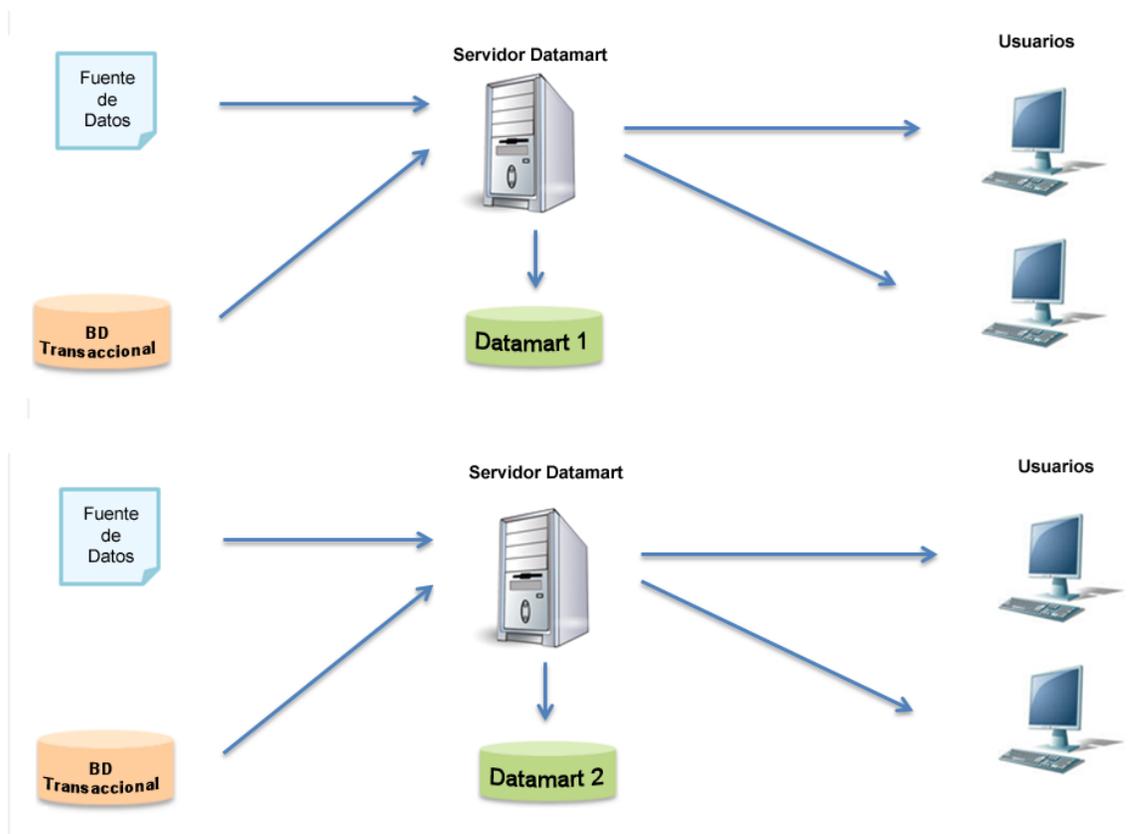
Figura 4. *Datamarts* dependientes



Fuente. Mannino, Michael. Administración de bases de datos diseño y desarrollo de aplicaciones. 2007. Pág. 557.[7]

Datamarts Independientes: No dependen de un *Data Warehouse* central, reciben los datos directamente de las bases de datos transaccionales u otras fuentes externas o internas.

Figura 5. *Datamarts* independientes



Fuente. Mannino, Michael. Administración de bases de datos diseño y desarrollo de aplicaciones. 2007. Pág. 558. [7]

2.5 OLTP Y OLAP

Existen dos tipos de sistemas importantes en el ámbito de *Data Warehouse*, como lo son OLTP y OLAP, en los siguientes párrafos se definen estas dos filosofías de procesamiento de la información:

2.5.1 On-line Transactional processing (OLTP). Constituye el trabajo primario en un sistema de información. Este trabajo consiste en realizar transacciones, es decir que puede involucrar operaciones de inserción, modificación o eliminación sobre las bases de datos transaccionales con un objetivo operacional: hacer funcionar las aplicaciones de la organización, proporcionar información sobre el estado del sistema de información y permitir actualizarlo conforme va variando la realidad del contexto de la organización. Muestras de este tipo de trabajo

transaccional son, por ejemplo, en el caso de una empresa, la inserción de un nuevo cliente, el cambio de sueldo de un empleado, la tramitación de un pedido, el almacenamiento de una venta, la impresión de una factura, la baja de un producto, etc. Es el trabajo diario y para el que inicialmente se ha diseñado la base de datos [3].

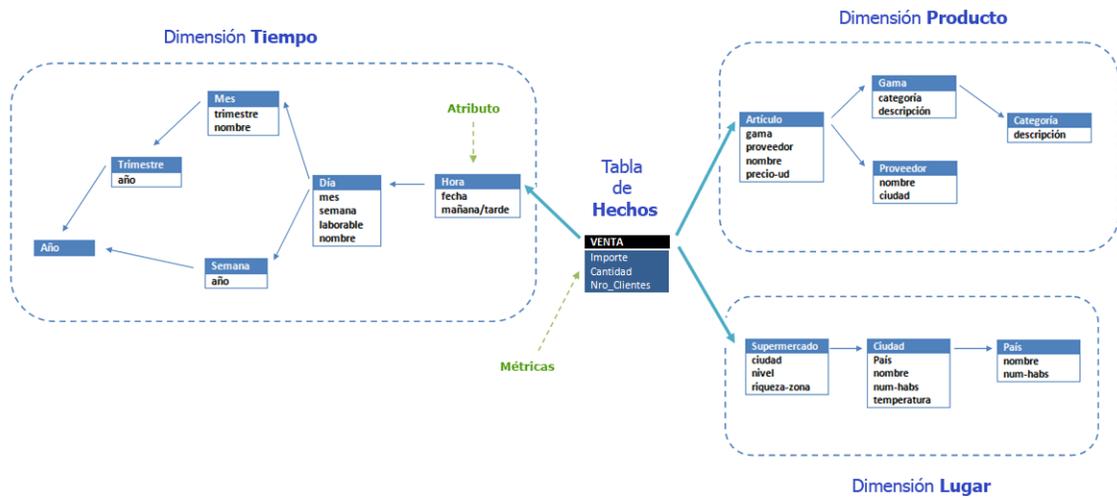
2.5.2 On-line Analytical Processing (OLAP).Engloba un conjunto de operaciones, exclusivamente de consulta en las que se requiere agregar y cruzar gran cantidad de datos e información. El objetivo de estas consultas es realizar informes y resúmenes, generalmente para el apoyo en la toma de decisiones. Ejemplos de este tipo de trabajo analítico pueden ser resúmenes de ventas mensuales, los consumos eléctricos por días, la espera media de los pacientes en cirugía digestiva en un hospital, el producto cuyas ventas han crecido más en el último trimestre, las llamadas por horas, etc. Este tipo de consultas suelen emanarse de los departamentos de dirección, logística o prospectiva y requieren muchos recursos. El acceso a los datos suele ser de sólo lectura, donde los datos se estructuran según las áreas de negocio, y los formatos de datos están integrados de una manera uniforme en toda la organización. Por lo general los sistemas *OLAP* se alimentan de información procedente de los sistemas de información operacionales existentes[11].

2.6 MODELO MULTIDIMENSIONAL

El modelo multidimensional es el modelo conceptual de datos más amplio para el desarrollo de *Data Warehouse*. Este modelo proporciona una interfaz intuitiva para los usuarios, permitiendo a estos trabajar de forma interactiva.

2.6.1 Estructura. El modelo multidimensional está compuesto por dos elementos, como se muestra en la figura 6, los cuales son medidas o tablas de hechos y dimensiones.

Figura 6. Estructura del modelo multidimensional



Fuente. Hernández, José y Ramírez, María. Introducción a la Minería de Datos. 2004. Pág. 49.[5]

Medidas o tablas de hechos. Contienen variables de medida y se asocian mediante punteros con las tablas de dimensión. Estas tablas de hechos contienen datos y los identificadores de las dimensiones de cada uno de los datos[3]. Por lo general se encuentran normalizadas y todos los atributos identificadores forman parte de la llave primaria, en donde cada identificador referencia a la dimensión correspondiente. Los atributos que representan la información de cada hecho se denominan métricas o medidas. Para las medidas se utilizan operaciones de agregación (promedio, conteo, suma, etc.), por lo cual las métricas a utilizar deben contener en lo posible valores numéricos, se debe hacer un análisis muy cuidadoso respecto a este tema para definir las operaciones a realizar, sobre el hecho que se quiere estudiar.

Dimensiones. Las dimensiones describen los hechos de manera general o detallada, por ejemplo si los hechos son ventas, las dimensiones podrían ser tiempo, ubicación, clientes y productos.

Las dimensiones pueden tener jerarquías, compuestas de un nivel como mínimo o de varios niveles. Por ejemplo. La dimensión *Location* puede tener una jerarquía compuesta por los niveles país, departamento y ciudad, en donde la granularidad más alta es país y la mínima es ciudad, o la jerarquía de la dimensión *Time* compuesta por año, trimestre, mes, semana, día y hora la granularidad más alta

es año y la mínima es hora[7].La granularidad es el nivel de detalle que posee cada registro de una tabla de hechos.

Cada dimensión tiene una estructura jerárquica pero no necesariamente lineal como se muestra en la Figura 6, en la dimensión tiempo, puede haber más de un camino de agregación (ruta de agregación). Esto permite diferentes niveles y caminos de agregación para las diferentes dimensiones, posibilitando la definición de hechos agregados con mucha facilidad [5].

Es evidente comprobar que las métricas en las tablas de hechos responden a valores cuantitativos, mientras que las dimensiones responden a valores cualitativos o de descripción.

Atributos. Proporcionan mayor nivel de detalle a las dimensiones, por los atributos se pueden encontrar jerarquías dentro de las dimensiones y propiedades de éstas. Los atributos pueden ser descriptivos tales como: “extra largo”, “número medio”, “sí”, “no”, entre otros.

2.6.2 Modelado conceptual de un *Data Warehouse*.

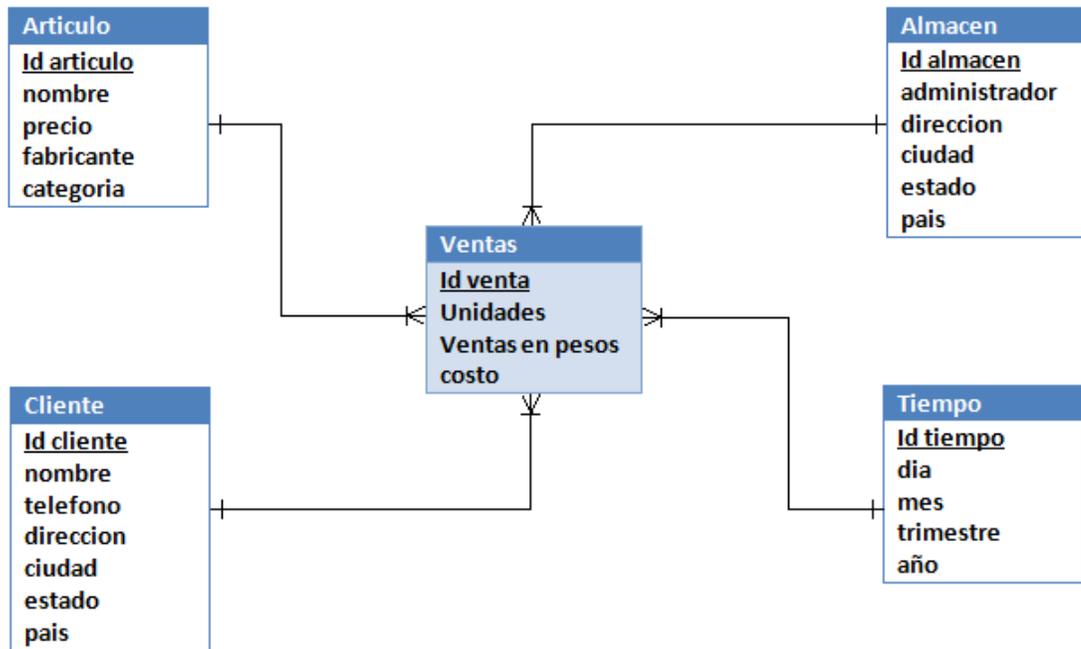
Esquema en estrella. Se basa en una tabla de hechos central conectado con un conjunto de tablas o relaciones de dimensiones y es una técnica de modelado que sirve para representar la información de manera multidimensional como se muestra en la Figura 7[7].A diferencia de la estructura de otros esquemas de base de datos.

Según Mannino [7] las características de este esquema son las siguientes:

- El modelo es fácil de entender para los usuarios.
- La llave primaria de la tabla de hechos representa a cada una de las dimensiones.
- Las columnas que no son llaves foráneas son valores o métricas en la tabla de hechos y valores descriptivos en las dimensiones.
- Las tablas de hechos están usualmente normalizadas.
- Las dimensiones están completamente desnormalizadas.
- Proveen una respuesta rápida a las consultas.
- El rendimiento de las consultas es mejorado reduciendo las uniones entre las tablas.

- Los usuarios pueden expresar consultas complejas.
- Se puede visualizar la información a través de múltiples herramientas.
- Fácil de definir jerarquías.
- Requiere bajo mantenimiento
- Provee flexibilidad para cambios y crecimientos de la información.

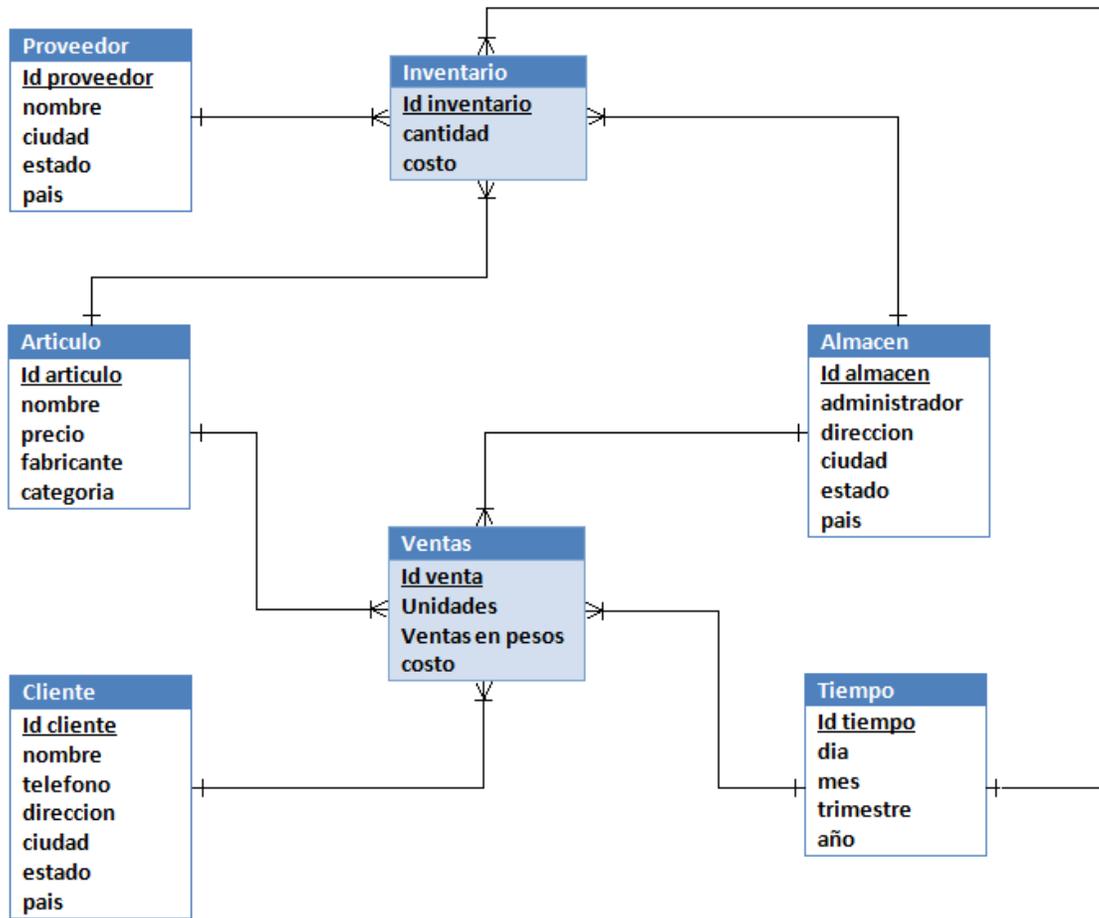
Figura 7. Esquema estrella



Fuente.Mannino, Michael. Administración de bases de datos diseño y desarrollo de aplicaciones. 2007. Pág. 568. [7]

Esquema de constelación. Este esquema es una extensión del esquema en estrella y está constituido por múltiples tablas de hechos en el centro relacionadas con tablas de dimensión. Las tablas de hechos por lo general comparten algunas tablas de dimensión como se muestra en la Figura 8.

Figura 8. Esquema de constelación



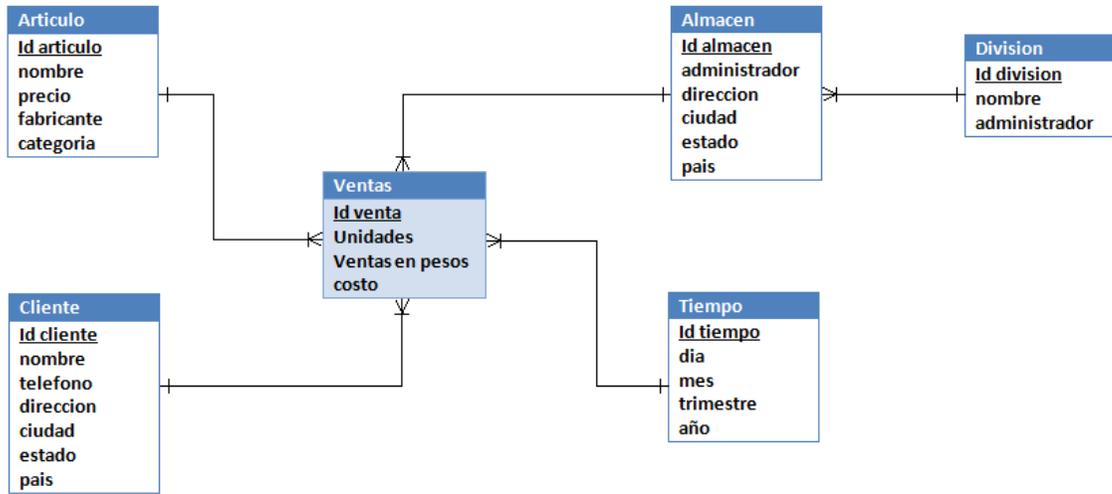
Fuente. Mannino, Michael. Administración de bases de datos diseño y desarrollo de aplicaciones. 2007. Pág. 569. [7]

Esquema de copo de nieve. Es una variante del esquema en estrella, donde algunas tablas de dimensiones están normalizadas, por lo cual tiene múltiples tablas de dimensiones en forma jerárquica como se muestra en la Figura 9.

Algunas características de este esquema son:

- Mejor utilización del espacio.
- Mayor complejidad en su estructura.
- Menor esfuerzo en el diseño, porque las tablas de dimensiones están normalizadas.
- Puede desarrollar clases jerárquicas fuera de las tablas de dimensiones, que permiten realizar análisis de lo general a lo más específico [12].

Figura 9. Esquema de copo de nieve

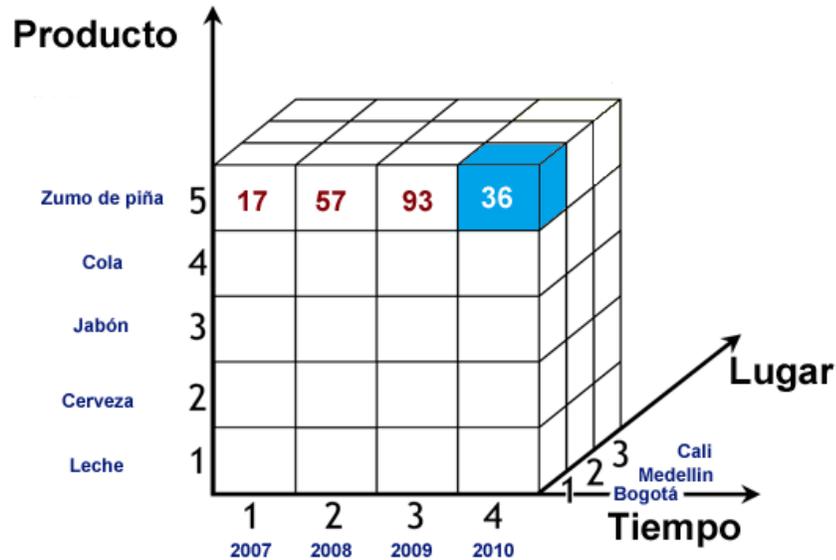


Fuente. Mannino, Michael. Administración de bases de datos diseño y desarrollo de aplicaciones. 2007. Pág. 570. [7]

2.6.3 Tecnologías de almacenamiento y optimización. Para entender las tecnologías es importante conocer el concepto de cubo de datos que se define a continuación.

2.6.3.1 Cubo de datos. Está formado por casillas, las cuales tienen medidas, por ejemplo valores numéricos como las cantidades de ventas unitarias. Cada casilla es un hecho que se describe a través de las dimensiones, para agrupar estos valores[7]. El cubo de datos mostrado en la Figura 10 da un ejemplo claro de un cubo multidimensional, aunque se puede conocer también como hipercubo si conceptualmente tiene más de tres dimensiones

Figura 10. Cubo de datos Multidimensional.



Fuente. Elmasri, Ramez y Navathe, Shamkant. Fundamentos de sistemas de bases de datos. 2007. Pág. 855. [3]

Los cubos, las dimensiones y atributos son la esencia de la navegación OLAP. Existen varias tecnologías de almacenamiento multidimensional. Cada una de ellas con diferentes características y aplican de acuerdo a las necesidades de cada organización. A continuación se presentan tres mecanismos de almacenamiento presentando algunas características, ventajas y desventajas.

2.6.3.2 MOLAP. Utiliza bases de datos multidimensionales para el análisis de los datos. Los datos se almacenan sobre una matriz optimizada. La información se almacena multidimensionalmente para ser visualizada de una forma resumida y proporcionando datos estadísticos. Utiliza una arquitectura de dos niveles: La base de datos multidimensional y el motor analítico. La base de datos multidimensional se encarga del manejo, acceso y obtención de los datos[4].

El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan el análisis OLAP[4].

Ventajas

- Consultas rápidas debido a la optimización del rendimiento de almacenamiento, la indexación multidimensional y la memoria cache.
- Ocupa menor tamaño en disco en comparación con los datos almacenados en bases de datos relacionales debido a técnicas de compresión.
- Automatización del procesamiento de los datos agregados de mayor nivel.
- Muy compacto para conjuntos específicos de pocas dimensiones.
- El modelo de almacenamiento en matrices proporciona una indexación natural.
- Eficaz extracción de datos lograda gracias a la pre-estructuración de los datos agregados.

Desventajas

- La etapa de procesamiento (carga de datos) puede ser bastante larga, sobre todo para grandes volúmenes de datos. Normalmente, este se puede evitar con un procesamiento incremental, es decir, sólo el procesamiento de los datos que han cambiado, en lugar de volver a procesar todo el conjunto de datos.
- Las herramientas MOLAP tradicionalmente tienen dificultades para consultar con modelos con dimensiones muy altas.
- El enfoque MOLAP introduce redundancia en los datos[7].
- Tienden a sufrir un bajo rendimiento cuando consultan elementos como descripciones textuales.

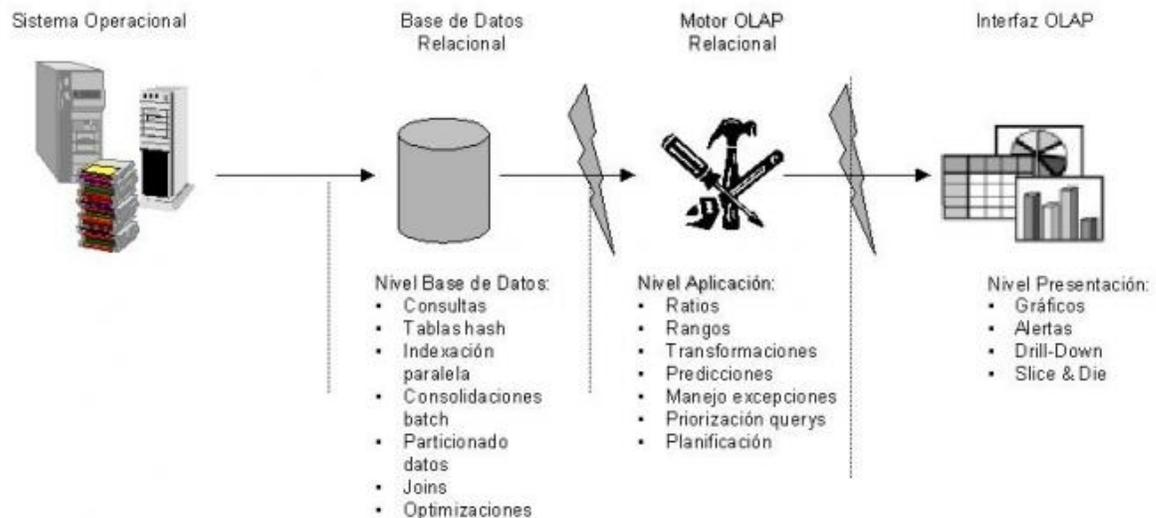
2.6.3.3 ROLAP. Debido al tamaño del mercado potencial y al uso de *Data Warehouse* en las organizaciones, se han creado productos con características adicionales para soportar operaciones y estructuras de almacenamiento para datos multidimensionales sobre sistemas de bases de datos relacionales, esto se conoce como ROLAP o también por OLAP relacional.

Las bases de datos relacionales utilizan los esquemas conceptuales de estrella, copo de nieve o de constelación, para almacenar los datos de manera multidimensional. Los cubos de datos se construyen dinámicamente a partir de tablas de hechos y dimensiones. Por lo regular, solo debe construirse un subconjunto de datos, como lo especifica la consulta de un usuario[7].

Este sistema utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención de los datos. El nivel de

aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios. El motor ROLAP se integra con los niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP.

Figura 11. Niveles ROLAP



Fuente. García Rodríguez, Byron. Sistema de análisis de información gerencial académico-financiero para F.I.S.C.C. (Data Mart). 2002. Pág. 26. [4]

Según Mannino[7], Dentro de las técnicas de almacenamiento se pueden nombrar las siguientes:

Reescritura de consultas usando vistas materializadas.

Pueden eliminar la necesidad de tener acceso a grandes tablas de hechos y dimensiones. Si las vistas materializadas son grandes, pueden indizarse para mejorar el desempeño de la recuperación. La reescritura de consultas utiliza el optimizador de consultas para evaluar el beneficio de emplear vistas materializadas en comparación con las tablas de hechos y dimensiones.

Asesores de almacenamiento de resumen.

Determina el mejor conjunto de vistas materializadas que debería crearse y mantenerse para una carga de trabajo de consulta dada. Para que haya consistencia con otros componentes, el asesor de resumen se integra con el componente de reescritura de consultas y el optimizador de consultas.

Partición, simplificación y ejecución paralela de consultas.

Ofrecen oportunidades para disminuir el tiempo de ejecución de las consultas al *Data Warehouse*. Es necesario estudiar con cuidado las elecciones, de modo que el uso de la partición y simplificación soporten el nivel deseado de ejecución paralela de consultas.

Ventajas

- Es más escalable para manejar grandes volúmenes de datos, especialmente modelos de dimensiones de gran cardinalidad.
- Los tiempos de carga son mucho menores que los de los sistemas MOLAP.
- Las herramientas para el análisis de la información no tienen que ser necesariamente OLAP, sino también se puede tener acceso mediante herramientas de informes SQL
- Es posible modelar datos con éxito que de otro modo no se ajustarían en un modelo dimensional estricto.

Desventajas

- El proceso de carga de tablas agregadas debe ser gestionado por el código ETL personalizado, las herramientas ROLAP no disponen de mecanismos automáticos para realizar esta tarea, lo que significa más tiempo de desarrollo de código.
- Existen funcionalidades propias de las herramientas MOLAP que no están disponibles en las herramientas ROLAP, como por ejemplo el indexado jerárquico especial.

2.6.3.4 HOLAP. El equilibrio entre MOLAP y ROLAP ha permitido desarrollar esta tercera tecnología, también conocida como OLAP híbrido, para combinar ROLAP y MOLAP. El HOLAP permite que un *Data Warehouse* se divida entre el almacenamiento relacional de tablas de hechos y dimensiones, y el almacenamiento multidimensional de cubos de datos de resumen. Cuando se presenta una consulta OLAP, el sistema HOLAP puede combinar los datos administrados por el ROLAP y MOLAP.

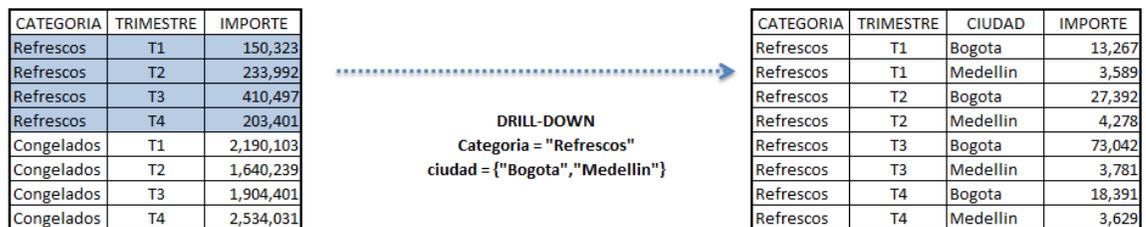
HOLAP tiene desventajas potenciales que pueden limitar su uso. Primero, HOLAP puede ser más complejo que ROLAP o MOLAP, en especial si un fabricante de DBMS no ofrece soporte total para HOLAP. Para poder soportar por completo HOLAP, un fabricante DBMS debe proporcionar mecanismos tanto MOLAP como ROLAP, así como herramientas para combinar ambos mecanismos en el diseño y operación de un *Data Warehouse*. Segundo, hay una transposición considerable

en la funcionalidad entre las técnicas de almacenamiento y la optimización ROLAP deben eliminarse o usarse además de las técnicas MOLAP. Tercero, debido a que la diferencia en el tiempo de respuesta se ha reducido entre ROLAP y MOLAP, su combinación quizá no ofrezca una mejora considerable en el desempeño como para justificar la complejidad mayor.[7].

2.6.4 Operaciones sobre el modelo multidimensional. La información que se analiza con OLAP debe estar estructurada de tal forma que se puedan realizar las siguientes operaciones:

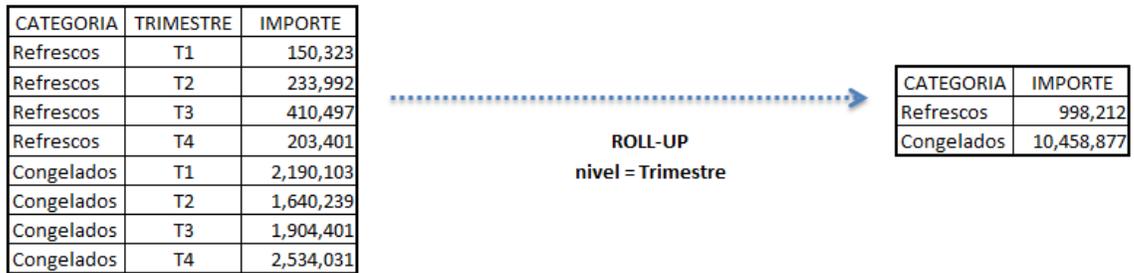
Drill Down y Roll Up (profundizar y escalar). Estas dos operaciones permiten visualizar la información a un nivel de detalle determinado. *Drill down* disgrega los datos (mayor nivel de detalle), partiendo de lo general a lo particular, por ejemplo, en la figura 12 se detalla información de refrescos, añadiendo atributos como la ciudad donde pertenecen. *Roll up* agrega los datos (mayor nivel de resumen), es decir de lo particular a lo general como en el ejemplo de la figura 13, en el cual la información de refrescos y congelados se observa detallada, para luego mostrarse en un resumen.

Figura 12. Drill Down



Fuente. Hernández, José y Ramírez, María Introducción a la Minería de Datos. 2004. Pág. 53.[5]

Figura 13. *Roll up*



Fuente. Hernández, José y Ramírez, María Introducción a la Minería de Datos. 2004. Pág. 54. [5]

Slice and dice. Estas dos operaciones permiten proyectar datos en el informe, navegando a través del cubo visualizado. La operación *Slice* corta el cubo para que el usuario pueda enfocarse solamente en algunas perspectivas, reduciendo la dimensionalidad del esquema eliminando alguna dimensión. La operación *Dice* hace que el cubo rote para poder apreciar la información desde otra perspectiva, es decir se restringen valores que se consideran en las dimensiones del cubo según alguna condición. En la operación *Dice* no se modifica la estructura del cubo de datos en cuanto a las dimensiones y niveles de estas, tampoco se modifica el nivel de detalle de los hechos pero si su número, dado que aquellos que tuvieran valores no considerados deben ser eliminados. En el ejemplo mostrado en la figura 14 se observa estas dos operaciones, mostrando únicamente los datos de los trimestres T1 y T4 discriminados por la categoría de refrescos y congelados en la ciudad de Bogotá [12].

Figura 14. *Slice & dice*

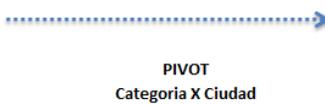


Fuente. Hernández, José y Ramírez, María Introducción a la Minería de Datos. 2004. Pág. 55. [5]

Pivot. Esta operación lo que persigue no es la modificación de la definición de la estructura del cubo de datos. En realidad lo que se realiza es un cambio en el orden de las dimensiones como se muestra en la figura 15 [12].

Figura 15. *Pivot*

CATEGORIA	TRIMESTRE	Bogota	Medellin
Refrescos	T1	13,267	3,589
Refrescos	T2	27,392	4,278
Refrescos	T3	73,042	3,781
Refrescos	T4	18,391	3,629
Congelados	T1	150,242	4,798
Congelados	T2	173,105	3,564
Congelados	T3	163,241	4,309
Congelados	T4	190,573	4,812



PIVOT
 Categoria X Ciudad

CATEGORIA	TRIMESTRE	Refrescos	Congelados
Bogota	T1	13,267	150,242
Bogota	T2	27,392	173,105
Bogota	T3	73,042	163,241
Bogota	T4	18,391	163,24
Medellin	T1	3,589	163,24
Medellin	T2	4,278	163,24
Medellin	T3	3,78	163,24
Medellin	T4	3,629	163,24

Fuente. Hernández, José y Ramírez, María Introducción a la Minería de Datos. 2004. Pág. 54.[5]

2.7 PROCESOS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE LOS DATOS (ETL)

Al crear un *Data Warehouse* implementarse mediante la tecnología ROLAP o MOLAP, el siguiente paso es la carga de datos. En las bases de datos el proceso similar al de carga se conoce como “migración”, aunque en la carga existe un mantenimiento posterior.

Este proceso es muy delicado y es el que demanda más esfuerzo, requiere aproximadamente la mitad del esfuerzo. Existe un sistema especializado para realizar esta tarea, denominado ETL (*extraction, transformation, load*). El sistema ETL es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformarlos y limpiarlos, y cargarlos en otra base de datos o *Data Warehouse* para el análisis de la información[12].

Este sistema no está disponible en el mercado, ni se consigue en internet, sino que su construcción es responsabilidad del equipo de desarrollo del *Data Warehouse* es específico para cada *Data Warehouse*. El ETL se puede construir realizando programas específicos o adoptando herramientas genéricas, como por ejemplo, el uso de disparadores o *trigger*, herramientas de migración o programas del mercado más específicos.

Según Timarán [12], entre las tareas del sistema ETL se puede nombrar las siguientes:

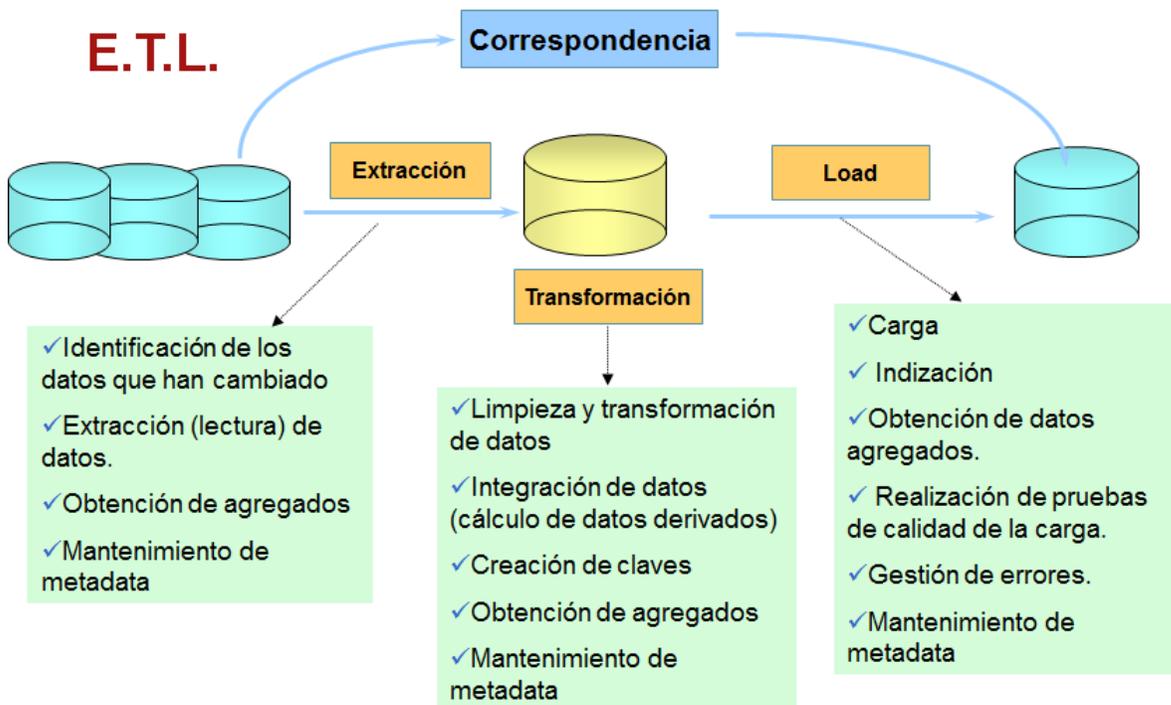
- Lectura de datos transaccionales.
- Incorporación de datos externos.
- Creación de claves o integridad referencial.

- Obtención de agregaciones.
- Limpieza y transformación de los datos.
- Creación y mantenimiento de los metadatos.
- Identificación de cambios.
- Planificación de la carga y mantenimiento.
- Indización.
- Pruebas de Calidad.

Los sistemas ETL realizan estas tareas mediante un repositorio de datos intermedio, y aunque parezca que se abusa de los recursos, este almacenamiento intermedio es extremadamente útil, debido a que hay tareas que no se pueden realizar en las bases de datos transaccionales o en el mismo *Data Warehouse*[5].

Las fases del sistema ETL se muestran en la figura 16.

Figura 16.Sistema ETL



Fuente.Timarán Pereira, Ricardo. Bodega de datos y procesamiento analítico en línea. 2009. Pág. 33. [12]

2.8 METODOLOGÍAS PARA UN PROYECTO DE *DATA WAREHOUSE*

En los siguientes párrafos se definen las diferentes metodologías para el diseño y construcción de *Data Warehouse*.

2.8.1 Metodología de Bill Inmon. Esta metodología es utilizada generalmente cuando la tecnología y los problemas del negocio están identificados claramente. Este enfoque logra la sinergia entre los problemas de negocio alcanzando los objetivos buscados. Se trata de un método sistémico, que minimiza los problemas de integración, pero es costoso, debido a la gran cantidad de datos y su poca flexibilidad. En este método se formula un resumen del sistema, sin especificar detalles. Cada parte del sistema se refina diseñándola con mayor detalle. Después, cada parte nueva se redefine, cada vez con mayor detalle, hasta que la especificación completa es lo suficientemente detallada para validar el modelo[9]. El modelo propone la validación de los requisitos una vez que se tiene el sistema.

Esta metodología la definió Bill Inmon en el año 1992 en el libro “*Building the Data Warehouse*”. En él proponía los mecanismos necesarios para llevar a cabo la correcta realización de un *Data Warehouse*[10].

Para Bill Inmon, el diseño de un *Data Warehouse* comienza con el proceso de carga de datos en el mismo, debido al gran volumen de datos que manipulan los *Data Warehouse* se hace necesario de un análisis cuidadosamente gestionado y condensado, de lo contrario dicho volumen de datos impide que los objetivos del *Data Warehouse* se alcancen.

Inmon sustenta uno de los principios fundamentales del desarrollo de un *Data Warehouse*, el principio que el ambiente de origen de los datos y el ambiente de acceso de datos deben estar físicamente separados en diferentes bases de datos y en equipos separados, también identifica la importancia de utilizar un *Data Warehouse* para guardar datos históricos continuos, ya que uno de los mayores obstáculos para el análisis de información relevante es no contar con datos disponibles sobre un periodo de tiempo extendido. Operacionalmente, se tiende a almacenar solamente una vista actual del negocio, lo cual es un período muy corto para un análisis serio de tendencias.

El *Data Warehouse* debe responder a las necesidades de todos los usuarios en la organización, y no sólo de un determinado grupo. Por lo general este enfoque se asocia frecuentemente al nivel empresarial, que involucran desde un inicio todo el

ámbito corporativo, sin centrarse en un área específica hasta después de haber terminado completamente el diseño del *Data Warehouse*. En su filosofía, un *Datamart* es sólo una de las capas del proyecto, los *Datamarts* son dependientes del *Data Warehouse* Corporativo y por lo tanto se construyen luego de él.

El enfoque desarrolla una estrategia inicial que consiste en la identificación de las áreas principales de la organización. Facilitando la construcción de una solución integral. Esto evita la aparición de situaciones inesperadas en el futuro cercano que puedan causar un retraso y poner en riesgo el desarrollo del proyecto, debido a que se conoce con antelación y exactitud la estructura que presentarán los principales núcleos del desarrollo, permitiendo centrar los esfuerzos del desarrollo para ser compatible con los sucesivos sistemas.

Por lo general el método utiliza la creación de una base de datos relacional con cierto nivel de normalización para el entorno en el que se implementará el *Data Warehouse* Corporativo. Inmon asegura que esta es la alternativa más adecuada con el que se logra una mayor eficiencia, sin afectar a los usuarios finales. Por otro lado aplica el esquema estrella o modelado dimensional a la aplicación *Front End* que se llama *Datamart*, y es donde realmente tiene lugar el acceso de los usuarios en su Arquitectura.

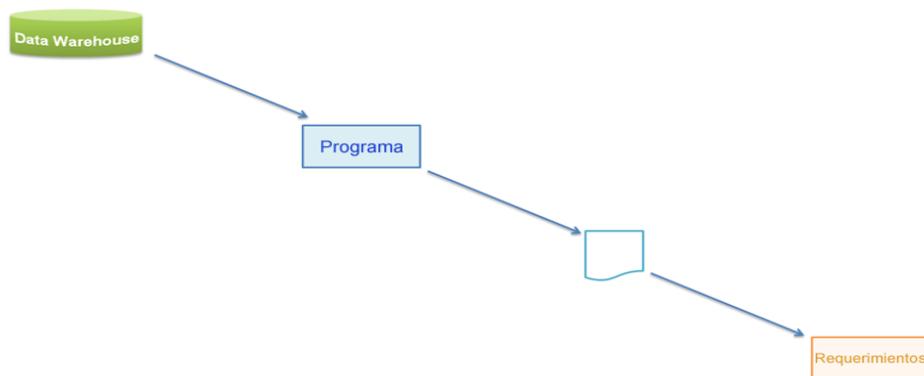
Por consecuencia Inmon está de acuerdo en que el modelado dimensional es adecuado para los *Datamart*, pero hace énfasis en que estos deben ser dependientes del *Data Warehouse* Corporativo; sin embargo está muy convencido que un diseño basado en Diagramas Entidad Relación E-R, es el más apropiado para un proyecto completo de *Data Warehouse*. Este diseño proporciona una estructura ideal y efectiva de coleccionar, almacenar y diseminar la Información, es muy probablemente que:

- *Datos antiguos*, limpiados en un RDBMS (potencialmente un *Data Warehouse* Empresarial).
- *Datos reconciliados*, desde el *Data Warehouse* Empresarial obtienen su información los *Datamarts*, cubos y otras herramientas para análisis y reportes que utilicen un enfoque multidimensional para mostrar la información.

La metodología tiene un enfoque en modo de “*Big Bang*” en el sentido de que en cierta forma no está relacionada al ciclo de vida normal de las aplicaciones, sino que los requisitos irán complementando al proyecto según vaya comprobándose su necesidad. Este enfoque es ideal para los propósitos de desarrollo del equipo

de Tecnología de Información pero no para las finanzas de una organización. Como esta estructura no es posible dividirla en partes modulares que comiencen a ser explotadas al implementarse, sino que es hasta que toda la arquitectura está en su lugar que los usuarios de negocio obtienen beneficio de ella. Este modo trae consigo un gran riesgo a la organización que invierte grandes esfuerzos en el desarrollo del *Data Warehouse* y solo hasta al final cuando aparecen los *Datamarts* donde realmente comienza a explotar su inversión y a obtener beneficios de ella.

Figura 17. Proceso del modelo Inmon



Fuente. Rodríguez Sanz, Miguel. Análisis y diseño de una *Data Mart* para el seguimiento académico de alumnos en un entorno universitario. 2010. pág. 25. [9]

La estrategia conlleva algunos riesgos, el más peligroso para un *Data Warehouse* es sin lugar a duda el caso cuando es imposible conocer en avance cuales son todas las necesidades de información de una empresa, el ambiente dinámico en que se mueve la organización, el cambio de estructura que conlleva el desarrollo de la nueva plataforma y los cambios resultantes a los sistemas operacionales que su introducción implica; es muy probable la posibilidad de que luego de un determinado tiempo y recursos invertidos en el desarrollo del *Data Warehouse*, una vez finalizado y puesto en funcionamiento, se hagan evidentes algunos cambios fundamentales que traen consigo altos costos de desarrollo para la organización, ocasionando que el éxito de todo el proyecto se ponga en riesgo, riesgos que podían ser evitados con una pronta detección en una temprana puesta en explotación de un primer avance del proyecto.

Por tal razón en esta metodología es frecuente a la hora de diseñar un sistema de información, utilizando las herramientas como el esquema Entidad Relación pero al tener un enfoque global, es más difícil de desarrollar en un proyecto sencillo,

pues estamos intentando abordar el “todo”, a partir del cual luego iremos al “detalle”[2]. Otra de las restricciones que trabajan en contra de esta metodología es que consume mucho más tiempo trabajar, lo que conlleva que muchas empresas se inclinen por usar metodologías de la que obtengan resultados tangibles en un espacio menor de tiempo.

2.8.2 Metodología de Ralph Kimball. La metodología para el desarrollo del proyecto será la creada por Ralph Kimball, quien es considerado una autoridad y uno de los padres de los *Data Warehouse*. La metodología de Kimball está orientada a que exista una mejor calidad en este tipo de proyectos.

Kimball ha establecido ciertos pasos para llevar al éxito un proyecto de *Data Warehouse*. Para su desarrollo se incluyen varias tareas que pueden ser realizadas en forma paralelo o en forma secuencial.

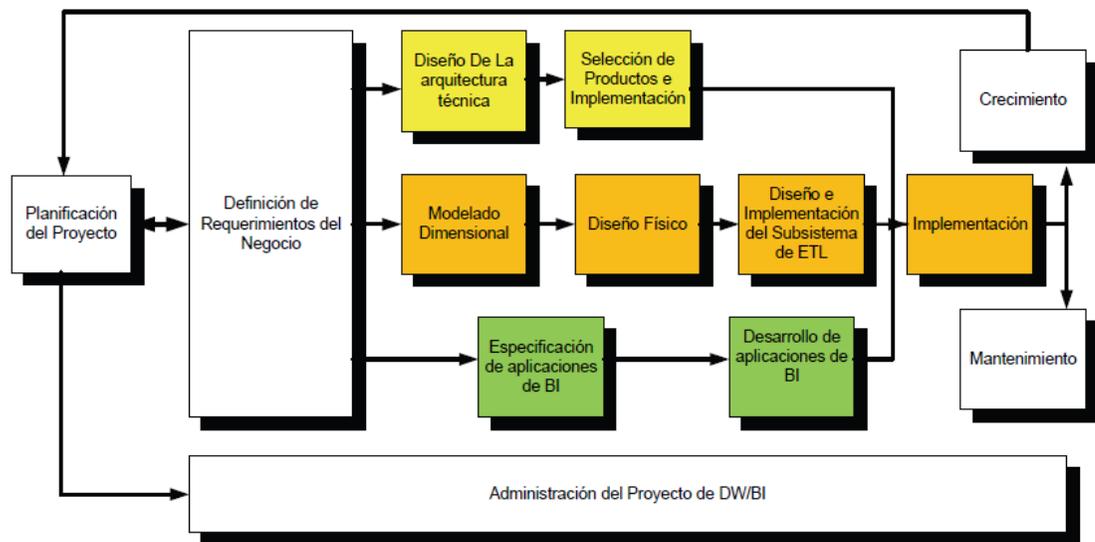
Con el desarrollo correcto de cada una de las fases de este modelo se garantiza la calidad y el correcto proceso de desarrollo.

Según Rivadera [8], la metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional de Negocio. Este ciclo de vida de *Data Warehouse*, está basado en cuatro principios básicos:

- **Centrarse en el negocio.** Aquí se resalta la identificación y la importancia de los requerimientos o las necesidades, con los cuales se desarrollará, mediante un gran esfuerzo, las relaciones consistentes con la organización, enfatizando el análisis de la misma.
- **Construir una infraestructura de información adecuada.** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la organización.
- **Realizar entregas en incrementos significativos.** Crear el *Data Warehouse* en incrementos entregables en plazos de 6 a 12 meses.
- **Ofrecer la solución completa.** Proveer todos los elementos necesarios para entregar valor a los usuarios de la organización. El *Data Warehouse* debe ser consistente, con un diseño adecuado, con calidad probada, y accesible. También se

deberá entregar herramientas de consulta *ad hoc*, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación. Las tareas (ciclo de vida) se muestran en la figura 18.

Figura 18. Tareas de la metodología Kimball.



Fuente. Rivadera, Gustavo. La metodología de Kimball para el diseño de almacenes de datos (*Data Warehouses*). 2010. Pág. 59. [8]

En la figura 18, se puede observar dos cuestiones. En primer lugar, hay que destacar la importancia de la definición de requerimientos, siendo los requerimientos de la organización el soporte inicial de las fases siguientes, además también tienen influencia en la fase de planeación del proyecto (nótese la doble flecha entre la caja de definición de requerimientos y la de planificación). En segundo lugar se puede observar tres caminos que se enfocan en tres diferentes áreas [8]:

- **Tecnología (Camino Superior).** Implica las tareas relacionadas con el software específico.
- **Datos (Camino del medio).** En la misma se diseñará el modelo dimensional y posteriormente el sistema ETL para la carga de datos en el *Data Warehouse*.

- **Herramientas de Inteligencia de Negocios (Camino Inferior).** Se encuentran las tareas de utilización o desarrollo de aplicaciones de análisis multidimensional para los usuarios finales.

A continuación se describen las fases de la metodología Kimball basados en la figura 18.

2.8.2.1 Planeación y administración del proyecto Existen varios posibles escenarios en los que se da un proyecto de *Data Warehouse* dentro de una organización, en donde se establece [8][12]:

- El objetivo general del proyecto.
- Los objetivos específicos.
- La delimitación y/o alcance del mismo.
- Programar las tareas necesarias para el desarrollo del proyecto.
- Planificar el uso de recursos.
- Asignar la carga de trabajo de los recursos.
- Elaboración de un documento final que representa un plan del proyecto[8].

Entre los diferentes escenarios se puede nombrar los siguientes:

- **Demanda de un sector de negocio.** En este escenario. Un ejecutivo del negocio tiene el propósito de obtener información con un mejor acceso para la toma de decisiones.
- **Demasiada demanda de información.** En este escenario, existen múltiples ejecutivos del negocio buscando mejor información.
- **En busca de demanda.** En este escenario, usualmente está involucrado el presidente de la organización, quien no identifica las necesidades de un *Data Warehouse* para su negocio, pero desea incorporar este sistema por razones diferentes a requerimientos o necesidades del negocio.

El identificar el escenario permite determinar si es necesario desarrollar el proyecto de *Data Warehouse* y determinar de dónde proviene la demanda de este. El primer caso es ideal, se cuentan con objetivos claros y con un alcance determinado de lo que se quiere en el proyecto. El segundo caso es riesgoso, implementar una bodega de datos que soporte varios requerimientos de diferentes

áreas de la empresa requiere de mucho tiempo, dinero y soporte interno de la organización a largo plazo. En el tercer caso se deben buscar los requerimientos que puedan implementar la solución y basar en ellos el proyecto.

2.8.2.2 Análisis de requerimientos. En este proceso se entrevista al personal de la organización. No se debe preguntar a estos usuarios, que datos quieren que aparezcan en el *Data Warehouse*, sino hablar con ellos acerca de su trabajo, objetivos y retos e intentar conocer cómo toman decisiones, actualmente y en el futuro[8][12].

Se debe considerar lo que necesita el negocio comparando estos requisitos con los datos almacenados en las bases de datos transaccionales que servirán como fuente, para lograr un soporte mucho mejor a estos requerimientos.

- **Preparación de la entrevista.** Se deben determinar roles y responsabilidades en el equipo que va a entrevistar a los usuarios y equipo de tecnología e informática de la organización preferiblemente.
- **Investigación previa a entrevistas.** Antes de iniciar el proceso de levantamiento de requerimientos, se deben analizar los reportes de la compañía, para determinar las decisiones y hechos estratégicos. Igualmente es ventajoso obtener los planes de negocio de la organización. Se debe analizar así mismo la competencia y las principales fortalezas y debilidades de la organización. Si ha existido proyectos de *Data Warehouse* esto también debe ser considerado.
- **Desarrollo del cuestionario.** Se deben desarrollar varios cuestionarios que serán aplicados dependiendo del rol de los entrevistados de la empresa. El cuestionario debe ser de una sola página para ahorrar el tiempo en las entrevistas.
- **Inicio y desarrollo de la entrevista.** La entrevista debe iniciarse con una introducción, para recordar al usuario sobre el proyecto y el equipo desarrollador. Los objetivos del proyecto y de la entrevista deben ser nombrados y los miembros del equipo deben ser debidamente presentados.

Para documentar información útil se debe preguntar a los usuarios sobre sus trabajos, por qué lo hacen y cómo los hacen. Se deben realizar preguntas en un alto nivel y luego al detalle para obtener respuestas cada vez más específicas.

- **Análisis de las entrevistas.** Si algún miembro del equipo conoce los sistemas operativos, debe explicarlos al resto del equipo para determinar la factibilidad de implementar los requerimientos encontrados. Se deben resaltar los descubrimientos y requerimientos claves para el proyecto.

Se deben analizar y repasar los reportes y análisis reunidos en las entrevistas, que comúnmente conlleva a una aproximación del descubrimiento de las dimensiones para el modelo.

Es importante documentar los requerimientos obtenidos y comunicarlos a los usuarios para adquirir su aprobación y compromiso.

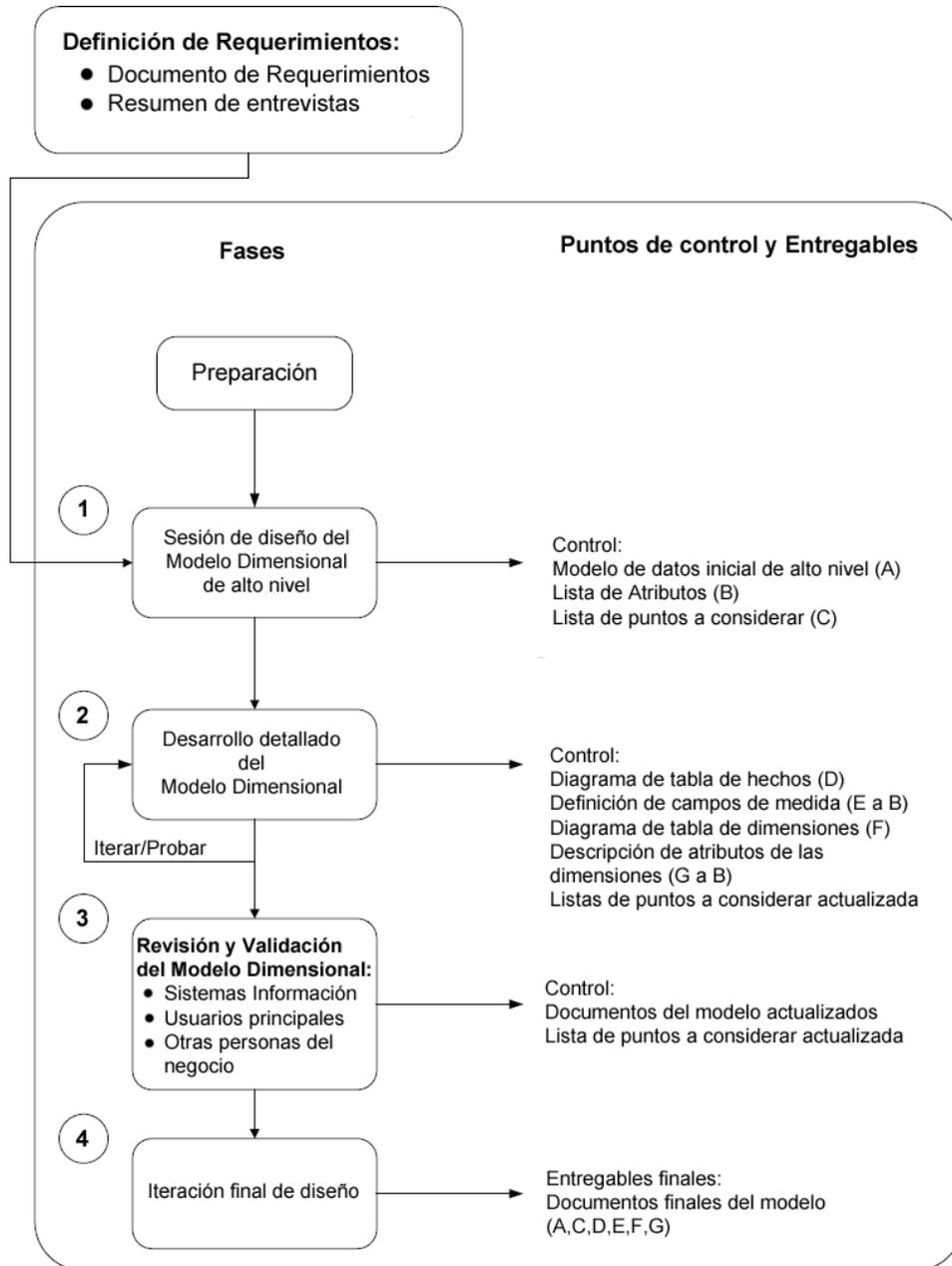
2.8.2.3 Modelado dimensional

- **Modelo entidad relación.** El modelo entidad relaciones una técnica poderosa para diseñar lógicamente el procesamiento de transacciones OLTP. Siempre va encaminado a la eliminación de la redundancia, lo que permite que la manipulación sobre la base de datos tenga que hacerse en un solo lugar y sea mucho más rápido ya que en el momento de la transacción requiera insertar, adicionar, borrar o modificar un dato es necesario que lo haga en un solo lugar y no en múltiples. Esto contribuye también al almacenamiento de grandes cantidades de datos dentro de las bases de datos relacionales, a través de este proceso de normalización. Por eso es perfecto para la inserción y actualización de la información.

Este es un modelo excelente para registrar las transacciones y administrar tareas operativas. Sin embargo, en el modelamiento de una bodega de datos presenta varios problemas. Los usuarios finales no entienden ni recuerdan un diagrama entidad relación. El uso del diagrama entidad relación va en contra del objetivo principal de una bodega de datos, de proporcionar datos de forma intuitiva y con buen desempeño y tiempos de respuesta.

- **Modelo Dimensional.** Una descripción más detallada de esta fase está descrita en la sección capítulo 2.6. Este es un proceso dinámico y altamente iterativo. Su esquema se representa en la Figura 19.

Figura 19. Proceso de Modelado Dimensional (Kimball)



Fuente. Rivadera, Gustavo. La metodología de Kimball para el diseño de almacenes de datos (*Data Warehouses*). 2010. Pág. 63. [8]

El proceso iterativo consiste en cuatro pasos: Elegir el proceso de negocio, establecer el nivel de granularidad, elegir las dimensiones e identificar las medidas y las tablas de hechos.

Elegir el proceso de negocio. El primer paso es elegir el área a modelar. Esta es una decisión de la dirección, y depende fundamentalmente del análisis de requerimientos y de temas analíticos recogidos en entrevistas.

Establecer el nivel de granularidad. La elección de la granularidad depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales. La sugerencia general es comenzar a diseñar el *Data Warehouse* al mayor nivel de detalle posible. En caso contrario no sería posible la operación *drill-down* mencionada anteriormente.

Elegir las dimensiones. Las dimensiones surgen de las discusiones del equipo, y se facilitan al elegir el nivel de granularidad. Los atributos de las dimensiones por lo general son atributos que son posibles candidatos para ser encabezados en cualquier forma de visualización, unidimensional o multidimensional.

Identificar medidas y las tablas de hechos. Las medidas se vinculan con el nivel de granularidad y se encuentran en las tablas de hechos. Las medidas se establecen de acuerdo a los requerimientos identificados dentro de un proceso organizacional[8].

Diseño de las dimensiones y hechos. Al diseñar las tablas de hechos y dimensiones, la idea principal es que cada dato de la organización sea representado como un cubo. Es importante que las tablas de hechos y dimensiones cumplan con las mismas especificaciones, como su granularidad.

Todas las llaves del *Data Warehouse* deben ser llaves subrogadas, es decir no deben significar nada respecto a las características de su contenido ni a su fuente en los sistemas fuente. No se deben utilizar las llaves originales de un sistema fuente extraídas. Estas llaves subrogadas se manejan como enteros.

- **Diseño Técnico de la arquitectura.** Con la definición de la arquitectura se mejora la comunicación entre las diferentes áreas de del proyecto, la planeación del proyecto, y el mantenimiento del mismo.

Aspectos de la arquitectura. Para hacer el diseño de la arquitectura se debe comenzar analizando los sistemas *legacy* actuales, estos deben ser consistentes y deben manejar de forma correcta sus transacciones.

Tabla 3: Bases de la arquitectura de *Data Warehouse*.

NIVEL DE DETALLE	Datos (el qué)	Back Room (el cómo)	Front Room (el cómo)	Infraestructura (el dónde)
Auditoria para los requerimientos del negocio	¿Qué información se necesita para tomar mejores decisiones en el negocio?	¿Cómo se hará el proceso ETL y cómo se tendrán los datos disponibles para los usuarios?	¿Cuáles son los principales retos que enfrenta el negocio? ¿Cómo se requieren analizar los datos?	¿Qué niveles de <i>Data Warehouse</i> se requieren para tener éxito? ¿Cuáles existen actualmente?
Modelos y documento de Arquitectura	Modelo Dimensional: ¿Cuáles son las principales entidades que componen esta información y cómo se relacionan?	¿Qué características específicas se necesitarán para tener los datos en una forma útil en lugares apropiados? ¿Cuáles son los principales <i>Data Warehouse</i> y donde deben estar localizados?	¿Qué requerirán los usuarios para cargar la información de forma útil?, ¿Qué clases de análisis y reportes se deben proveer y cuáles son las prioridades?	¿Cuál es el origen y el destino de los datos?, ¿Se tiene suficiente capacidades procesamiento y almacenamiento?
Especificaciones y modelos	Modelos lógicos y físicos: ¿Cuáles son los elementos individuales, sus derivaciones y reglas para la derivación?, ¿Cuáles son los recursos y	¿Qué estándares y productos proveen las características requeridas?, ¿Cuáles son los estándares de desarrollo,	¿Cuáles son las especificaciones de reportes incluyendo filas, columnas, secciones, encabezados, etc.?, ¿Quién las necesita, qué tan a	¿Cómo se interactúa con estas características? ¿Qué utilidades existen en el sistema, API's, etc.?

Tabla 3. Continuación.

NIVEL DE DETALLE	Datos (el qué)	Back Room (el cómo)	Front Room (el cómo)	Infraestructura (el dónde)
	cómo mapear a los destinos?	administración de código y nombres?	menudo?, ¿Cómo se distribuirán?	
Implementación	¿Crear bases de datos índices, <i>Back up</i> ?	Escritura de extracciones y cargas, automatización de procesos, documentación	Implementación del ambiente del reporte y análisis, Construcción del conjunto de reportes inicial, entrenamiento a usuario, documentación	Instalación y pruebas de nuevos componentes de la infraestructura. Conexión de las fuentes a los destinos y los clientes. Documentación.

Fuente. Timarán Pereira, Ricardo. Bodega de datos y procesamiento analítico en línea. 2009. [12]

En la tabla 3 se muestran las bases para la estructura de la arquitectura de *Data Warehouse*. Entre las principales se pueden destacar los datos, el área técnica y la infraestructura. Estas áreas de la arquitectura son interceptadas por filas que nos indican el detalle del incremento.

Definición de columnas

Datos. El contenido de un *Data Warehouse* son datos. Estos se refieren a todo el contenido físico de este el cual permite el análisis multidimensional y los reportes para el apoyo a la toma de decisiones. Los datos guardan toda la información del ambiente del *Data Warehouse* y de las fuentes que proveen a este.

Técnica. Esta área corresponde a los procesos y a las herramientas que se aplican sobre los datos, respondiendo a las preguntas “*Cómo*”. Por ejemplo ¿Cómo se extraerán los datos de las fuentes?, ¿Cómo se puede organizar los datos de forma que se pueda hacer un análisis consistente del negocio?, etc. Este tema se enfoca principalmente a las herramientas, al desarrollo del *Data Warehouse* y al mantenimiento del mismo. Para garantizar un mantenimiento del área técnica esta se divide en:

- Back Room: Es el área del *Data Warehouse* responsable de extraer y preparar los datos.
- Front Room: Es el área responsable de *Data Warehouse* de mostrarle a los usuarios los resultados con los datos analizados y obtenidos, listos para que puedan ser utilizados.

Infraestructura. Corresponden a las plataformas sobre las que se ejecutan los servidores de bases de datos, los servidores de aplicaciones y donde se ejecutan los procesos. La infraestructura es la planta física del *Data Warehouse*, se refiere principalmente al hardware utilizado para el desarrollo del proyecto.

Definición de los niveles de detalle. Se comenzará desde los niveles más altos de detalle. Parte de esta arquitectura trata de hacer modelos y documentos de diferentes niveles de detalle, algunos cercanos y otros no tan cercanos a la realidad.

Los mayores niveles de detalle de la arquitectura se usan para diseñar una estructura y obtener un mejor entendimiento de los niveles de detalle más bajos. Los modelos detallados ayudan a tomar rápidamente las especificaciones que se deben tener en cuenta durante el diseño y que por ningún motivo se pueden pasar por alto, pues de esta forma no se tendría un entendimiento total del negocio.

Nivel de requerimientos del negocio. Este nivel no trata ninguna implementación técnica del *Data Warehouse*, el interés del arquitecto del proyecto se centra en entender el comportamiento de la organización, sus procesos y las limitaciones que podrían ir en contra del desarrollo del *Data Warehouse*.

Nivel de modelos de la arquitectura. En este nivel se piensa en responder a los requerimientos. Un modelo de la arquitectura, propone los principales componentes de una arquitectura que se deben implantar para consecución de los requerimientos. Todas las tecnologías de arquitectura deben ser justificadas y deben garantizar que funcionan juntas en un sistema. Se deben definir si los componentes técnicos se comunican unos con otros, si las suposiciones administrativas para usar la tecnología son razonables y si la organización tiene recursos para soportar esta tecnología.

Nivel de detalle del modelo. Hace referencia a las especificaciones funcionales de cada componente de la arquitectura, esto debe incluir suficiente información que sirva como guía del equipo de desarrolladores para tener una implementación más confiable y con menos errores. Esto también sirve también en el caso de que se quiera establecer un contrato legal, pues los establecidos en la especificación

es lo mismo que se va a implementar y se puede evitar que el cliente exija más funcionalidades cuando nunca fueron identificadas formalmente.

El modelo dimensional y el modelo físico de una tabla de hechos corresponden al modelo de detalle para el área de datos.

Nivel de implementación. La implementación se realiza a partir de los detalles del modelo, pues ahí se tiene considerado la arquitectura y detalles para llevar a cabo el desarrollo. El desarrollo del proyecto debe estar documentado.

2.8.2.4 Procesos de extracción, transformación y carga. Los fundamentos teóricos de este proceso fueron definidos al final del capítulo anterior. Debido a que en este proyecto se desarrolla hasta la fase de diseño, dará una breve descripción de los pasos a seguir en esta fase ya que no hace parte del diseño.

- **Paso 1. Plan de alto nivel.** El proceso de diseño se inicia con un simple esquema de los componentes de plan que son conocidos: Las fuentes y los destinos de los datos, identificando de donde provienen, sus características y los problemas con estas fuentes. Este paso sirve para comunicar la complejidad del proyecto a la gerencia y miembros del equipo de desarrollo del proyecto.
- **Paso 2. Herramientas ETL.** Existen herramientas que realizan todo el proceso de extracción, transformación y carga que buscan minimizar el tiempo requerido para estas tareas. Estas herramientas implican un costo ya sea por licencias o compatibilidades u otros factores.
- **Paso 3. Plan detallado.** Se seleccionan las tablas con las que se va a trabajar, con los diferentes diagramas de estas estructuras. Las tablas de dimensión deben ser cargadas antes que las tablas de hechos.
- **Paso 4. Poblar una tabla de dimensión simple.** La razón para poblar una tabla de dimensión simple es la facilidad para migrar los datos hacia esta. Entre los pasos que se realizan están los siguientes: Extracción de la dimensión, Transformación de una dimensión y por último la carga de los datos.
- **Paso 5. Implementación de la lógica de cambio de una dimensión.** Se deben extraer únicamente los datos que han cambiado. El primer paso para preparar la dimensión consiste en determinar si ya existe un registro antiguo. Al encontrar un cambio se aplica alguno de los siguientes tipos de cambio:

Tipo 1: Sobrescribir. Se toma el nuevo dato de la fuente y se sobrescribe la tabla de dimensión con el nuevo contenido.

Tipo 2: Crear un nuevo registro de la dimensión. Se crea un nuevo registro de la dimensión con una nueva llave subrogada.

Tipo 3. Bajar el antiguo valor. Se crea un nuevo campo en la dimensión que contenga el antiguo valor del dato modificado y se actualiza la dimensión con el nuevo contenido.

- **Paso 6. Poblar las dimensiones restantes.** Para poblar el resto de las dimensiones se sigue el proceso del paso 4.

- **Paso 7. Carga Histórica de Hechos.** En el proceso ETL debe existir un paso para reemplazar las llaves primarias de las fuentes por las llaves subrogadas que se han asignado a cada dimensión, y que deben ir como llaves foráneas en la tabla de hechos. Este proceso funciona bien para bodegas de datos que no manejan grandes cantidades de registros, pues al ser grande el tamaño de los datos es preferible mantener una tabla de búsqueda que registre la llave primaria en la fuente y la llave subrogada en la dimensión. Una vez se cuenta con todas las llaves foráneas de la tabla de hechos la tabla se empieza a probar.

- **Paso 8. ETL de una tabla de hechos incremental.** Al realizar cargas semanales o periódicas desde las fuentes (actualizar el *Data Warehouse*), se deben procesar únicamente las transacciones que han ocurrido luego de la última carga.

- **Paso 9. Operación y automatización de un *Data Warehouse*.** Las operaciones típicas que se deben tener en cuenta en este paso son: Definición de tarea, horarios de tareas, monitoreo, *logging*, Manejo de Excepciones, Manejo de errores y notificaciones. Entre los procesos automáticos que se realizan se encuentran la Extracción y la calidad y limpieza de los datos.

También se debe considerar el mejoramiento de los datos, para garantizar el uso de los mejores datos posibles para el *Data Warehouse*. Además también se debe asegurar la calidad para verificar que el contenido es realmente correcto.

2.8.2.5 Selección e instalación de los productos. En esta fase se tienen en cuenta los siguientes puntos:

- **Selección de productos.** Una vez se tiene completa la arquitectura, se deben tener en cuenta dos componentes fundamentales para la selección del producto: Requerimientos técnicos y requerimientos del negocio.

Para la selección de la herramienta se debe hacer un análisis riguroso de cada una de ellas. Además siempre se debe tener en cuenta cuales son los requerimientos de los usuarios y del negocio, teniendo en cuenta las limitaciones del *hardware*, entre otros puntos importantes.

- **Características de aplicaciones para usuarios finales.** En los pasos explicados anteriormente se ha analizado el diseño e implementación; ahora se profundizará más en el *front room*. Este proporciona la interfaz que mostrará al usuario reportes y análisis multidimensionales que se tomará como base en la toma de decisiones.

Una aplicación de usuario final, provee un diseño y estructura a los reportes tomando como base los datos del *Data Warehouse*.

2.8.2.6 Mantenimiento y crecimiento de un Data Warehouse. En un proyecto de *Data Warehouse* se debe pensar en el mantenimiento posterior a la implementación, pues estas soluciones de inteligencia de negocios tienen gran tendencia a crecer a medida que crece la información dentro de la organización. Es bueno aclarar que aunque la inversión en el *Data Warehouse* es bastante importante, esta herramienta retorna la inversión que se le hace.

3. DISEÑO DEL DATAMART DE INFORMACIÓN ACADÉMICA

Para el diseño del *Datamart* se sigue la metodología estudiada de Ralph Kimball para la construcción de *Data Warehouse*, dado que establece claros procesos para todo el ciclo de desarrollo del proyecto, garantizando la calidad y eficiencia de la solución de inteligencia de negocios hasta la fase de diseño y es adaptable en la construcción de *Datamarts*.

Esta metodología fue aplicada desde el inicio del proyecto, hasta la fase de diseño y documentación del proyecto. En los siguientes puntos se describen los procesos realizados para cada fase del proyecto que garantizan su calidad y cumplimiento.

3.1 PLANEACIÓN Y ADMINISTRACIÓN DEL PROYECTO

3.1.1 Objetivo del proyecto. Diseñar una solución de Inteligencia de Negocios para soportar la toma de decisiones de los directivos de la Universidad de Nariño de una manera acertada, mediante un *Datamart* que integre la información de las bases de datos de la Oficina de Registro de Control Académico y el Centro de Informática.

3.1.2 Definición del proyecto. Este proyecto surge al identificar una alta demanda de información académica de los directivos de la Universidad de Nariño. La información se ubica en diferentes dependencias y la mejor manera de satisfacer estas necesidades es mediante una solución de Inteligencia de Negocios, en este caso el *Datamart*, el cual permite tratar la información de una manera específica (en este caso el área académica), histórica, integrada y a través de la línea del tiempo.

3.1.3 Alcance del Proyecto. El proyecto contiene los componentes de un *Datamart* hasta la fase de diseño. Como fuentes se tienen las bases de datos de información académica de la Oficina de Registro de Control Académico. En el proceso se construyó un prototipo funcional implementado en una base de datos en *PostgreSQL* denominada *datamartudenar*. El análisis multidimensional se realiza mediante la herramienta *STARCUBE1.0*, desarrollada en el laboratorio KDD del grupo de investigación GRIAS del programa de Ingeniería de Sistemas de la Universidad de Nariño, bajo los sistemas operativos Microsoft Windows y Linux.

El diseño de *Datamart* está planteado para soportar información académica acerca de inscripciones, admisiones, matrículas, rendimiento académico de estudiantes e información de profesores. Se espera en el futuro una total implementación de esta solución de inteligencia de negocios que de soporte a la toma de decisiones a los directivos de la Universidad de Nariño en los temas académicos.

3.1.4 Justificación del proyecto. Además de ser un proyecto académico, se busca fortalecer los procesos académicos referentes a la toma de decisiones en la Universidad de Nariño, proporcionando información fácil de entender, disponible, resumida e histórica en la línea del tiempo.

3.2 ANÁLISIS DE REQUERIMIENTOS

Para el levantamiento de requerimientos se realizaron entrevistas a los funcionarios del Centro de Informática, además se analizaron documentos también pertenecientes a esta área. Con la información recolectada en este proceso se realizó el modelo multidimensional que se describirá en la sección 3.3. A continuación se describe los requerimientos obtenidos.

3.2.1 Documentación de requerimientos. La documentación de los requerimientos pretende presentar una descripción de cada una de las fuentes de donde se extrajeron los datos para producir la información de la cual cada caso es responsable. Los requerimientos se clasificarán según cada área: inscripciones y admisiones, matrículas y población estudiantil, materias y rendimiento académico, y profesores.

3.2.1.1 Área de inscripciones y admisiones. En este proceso se tiene en cuenta información de aspirantes, teniendo en cuenta datos como el número de aspirantes, junto con sus datos personales, datos de inscripciones realizadas, además de información de las carreras a las que aspira. En la tabla 4 se describen los requerimientos obtenidos en el área de inscripciones y admisiones.

Tabla 4. Requerimientos inscripciones y admisiones.

Ref. #	Descripción	Prioridad
R 1.1	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por programa.	Alta
R 1.2	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por facultad.	Alta

Tabla 4. Continuación

Ref. #	Descripción	Prioridad
R 1.3	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por departamento	Alta
R 1.4	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por sede	Alta
R 1.5	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por año.	Alta
R 1.6	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por semestre.	Alta
R 1.7	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por ciudad.	Alta
R 1.8	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por sexo.	Alta
R 1.9	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por estrato.	Alta
R 1.10	El sistema deberá permitir conocer el número de aspirantes y/o admitidos por colegio.	Normal
R 1.11	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por programa	Alta
R 1.12	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por facultad	Alta
R 1.13	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por departamento	Alta
R 1.14	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por sede	Alta
R 1.15	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por año	Alta
R 1.16	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por semestre	Alta
R 1.17	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por ciudad	Alta
R 1.18	El sistema deberá permitir conocer el puntaje promedio de los aspirantes y/o admitidos por sexo	Alta

Tabla 4. Continuación

Ref. #	Descripción	Prioridad
R 1.19	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por programa	Normal
R 1.20	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por facultad	Normal
R 1.21	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por departamento	Normal
R 1.22	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por sede	Normal
R 1.23	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por año	Normal
R 1.24	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por semestre	Normal
R 1.25	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por ciudad	Normal
R 1.26	El sistema deberá permitir conocer el promedio de edad de los aspirantes y/o admitidos por sexo	Normal
R 1.27	El sistema deberá permitir conocer la relación entre aspirantes y admitidos	Normal

Fuente. Esta investigación.

3.2.1.2 Área de matrículas y población estudiantil. En este proceso se tiene en cuenta información de estudiantes matriculados y no matriculados, y población estudiantil, teniendo en cuenta datos como el número de estudiantes matriculados, junto con sus datos personales, información de las carreras a las que se matriculó o carreras a las que pertenece. En la tabla 5 se definen los requerimientos obtenidos en el área matrículas de estudiantes.

Tabla 5. Requerimientos matrículas e información de estudiantes

Ref. #	Descripción	Prioridad
R 2.1	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por programa	Alta
R 2.2	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por facultad.	Alta
R 2.3	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por departamento	Alta
R 2.4	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por sede	Alta
R 2.5	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por año.	Alta
R 2.6	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por semestre.	Alta
R 2.7	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por ciudad.	Alta
R 2.8	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por sexo.	Alta
R 2.9	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados clasificados por estado civil	Normal
R 2.10	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por estrato.	Alta
R 2.11	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por colegio.	Baja
R 2.12	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por tipo de cupo	Alta
R 2.13	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados discapacitados	Baja

Tabla 5. Continuación

Ref. #	Descripción	Prioridad
	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados con capacidades especiales	Baja
R 2.15	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados por tipo de etnia	Normal
R 2.16	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados clasificados por estado del estudiante (No graduado, Egresado y Graduado)	Alta
R 2.17	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados de primer semestre	Alta
R 2.18	El sistema deberá permitir conocer el número de estudiantes matriculados y/o no matriculados clasificados por edad	Normal
R 2.19	El sistema deberá permitir conocer el número de estudiantes por programa	Alta
R 2.20	El sistema deberá permitir conocer el número de estudiantes por facultad	Alta
R 2.21	El sistema deberá permitir conocer el número de estudiantes por departamento	Alta
R 2.22	El sistema deberá permitir conocer el número de estudiantes por sede	Alta
R 2.23	El sistema deberá permitir conocer el número de estudiantes por año	Alta
R 2.24	El sistema deberá permitir conocer el número de estudiantes por semestre	Alta
R 2.25	El sistema deberá permitir conocer el número de estudiantes por ciudad	Alta
R 2.26	El sistema deberá permitir conocer el número de estudiantes por sexo	Alta
R 2.27	El sistema deberá permitir conocer el número de estudiantes clasificados por estado civil	Alta
R 2.28	El sistema deberá permitir conocer el número de estudiantes por estrato	Alta
R 2.29	El sistema deberá permitir conocer el número de estudiantes por colegio	Alta
R 2.30	El sistema deberá permitir conocer el número de estudiantes por tipo de cupo	Alta

Tabla 5. Continuación

Ref. #	Descripción	Prioridad
R 2.31	El sistema deberá permitir conocer el número de estudiantes discapacitados	Normal
R 2.32	El sistema deberá permitir conocer el número de estudiantes con capacidades especiales	Normal
R 2.33	El sistema deberá permitir conocer el número de estudiantes por tipo de etnia	Baja
R 2.34	El sistema deberá permitir conocer el número de estudiantes por edad de ingreso	Normal
R 2.35	El sistema deberá permitir conocer la edad promedio de los estudiantes	Alta
R 2.36	El sistema deberá permitir conocer el promedio de edad o fecha de ingreso de los estudiantes	Alta
R 2.37	El sistema deberá permitir conocer el número de estudiantes clasificados por estado actual	Alta
R 2.38	El sistema deberá permitir conocer el número de estudiantes clasificados por el tipo de sangre	Baja
R 2.39	El sistema deberá permitir conocer el número de estudiantes clasificados por estado del estudiante (No graduado, Egresado y Graduado)	Alta

Fuente. Esta investigación.

3.2.1.3 Área de materias y rendimiento académico. En este proceso se tiene en cuenta información de materias matriculadas y canceladas por los estudiantes, además de la información de notas acumuladas en el tiempo, teniendo en cuenta datos de las carreras a las que se matriculó las materias, además generalizando la información personal de estudiantes, entre otra información. En la tabla 6 se definen los requerimientos obtenidos en el procesomatrículas de materias.

Tabla 6. Requerimientos de matrículas de materias

Ref. #	Descripción	Prioridad
R 3.1	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por programa	Alta
R 3.2	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por facultad.	Alta
R 3.3	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por departamento	Alta

Tabla 6. Continuación

Ref. #	Descripción	Prioridad
R 3.4	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por sede	Alta
R 3.5	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por año.	Alta
R 3.6	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por semestre.	Alta
R 3.7	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por ciudad.	Baja
R 3.8	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por sexo del estudiante.	Normal
R 3.9	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas por estrato.	Normal
R 3.10	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas tipo de cupo	Baja
R 3.11	El sistema deberá permitir conocer el número de materias matriculadas y/o canceladas clasificadas por tipo de materia	Normal
R 3.12	El sistema deberá permitir conocer el número de materias matriculadas y/canceladas clasificadas por estado del estudiante	Normal

Fuente. Esta investigación.

En la tabla 7 se definen los requerimientos obtenidos en el proceso rendimiento académico.

Tabla 7. Requerimientos de rendimiento académico

Ref. #	Descripción	Prioridad
R 3.1	El sistema deberá permitir conocer el promedio de notas definitivas por programa	Alta
R 3.2	El sistema deberá permitir conocer el promedio de notas definitivas por facultad.	Alta
R 3.3	El sistema deberá permitir conocer el promedio de notas definitivas por departamento	Alta
R 3.5	El sistema deberá permitir conocer el promedio de notas definitivas por sede	Alta
R 3.6	El sistema deberá permitir conocer el promedio de notas definitivas por año.	Alta

Tabla 7. Continuación

Ref. #	Descripción	Prioridad
R 3.7	El sistema deberá permitir conocer el promedio de notas definitivas por semestre.	Alta
R 3.8	El sistema deberá permitir conocer el promedio de notas definitivas por ciudad.	Alta
R 3.9	El sistema deberá permitir conocer el promedio de notas definitivas sexo del estudiante.	Normal
R 3.10	El sistema deberá permitir conocer el promedio de notas definitivas por estrato.	Alta
R 3.12	El sistema deberá permitir conocer el promedio de notas por tipo de cupo del estudiante	Normal
R 3.13	El sistema deberá permitir conocer el promedio de notas clasificadas por tipo de materia	Normal
R 3.14	El sistema deberá permitir conocer el promedio de notas clasificadas por estado del estudiante (No graduado, egresado, no egresado)	Normal

Fuente. Esta investigación.

3.2.1.4 Área de profesores. En este proceso se tiene en cuenta información de profesores y evaluación docente, teniendo en cuenta la información personal de los docentes y de las vinculaciones de los mismos, detallando el nivel académico, además de tener en cuenta los diferentes criterios de evaluación docente. En la tabla 8 se definen los requerimientos obtenidos en el área de profesores.

Tabla 8. Requerimientos de profesores

Ref. #	Descripción	Prioridad
R 4.1	El sistema deberá permitir conocer el número de docentes hora cátedra por facultad	Normal
R 4.2	El sistema deberá permitir conocer el número de docentes hora cátedra por genero	Normal
R 4.3	El sistema deberá permitir conocer el número de docentes hora cátedra por año	Normal
R 4.4	El sistema deberá permitir conocer el número de docentes tiempo completo por facultad	Normal
R 4.5	El sistema deberá permitir conocer el número de docentes tiempo completo por genero	Normal
R 4.6	El sistema deberá permitir conocer el número de docentes tiempo completo por año	Normal

Tabla 8. Continuación

Ref. #	Descripción	Prioridad
R 4.7	El sistema deberá permitir conocer el número de docentes con especialización por facultad	Alta
R 4.8	El sistema deberá permitir conocer el número de docentes con especialización por genero	Alta
R 4.9	El sistema deberá permitir conocer el número de docentes con especialización por año	Alta
R 4.10	El sistema deberá permitir conocer el número de docentes con maestría por facultad	Alta
R 4.11	El sistema deberá permitir conocer el número de docentes con maestría por genero	Alta
R 4.12	El sistema deberá permitir conocer el número de docentes con maestría por año	Alta
R 4.13	El sistema deberá permitir conocer el número de docentes con doctorado por facultad	Alta
R 4.14	El sistema deberá permitir conocer el número de docentes con doctorado por genero	Alta
R 4.15	El sistema deberá permitir conocer el número de docentes con doctorado por año	Alta

Fuente. Esta Investigación

3.3 MODELADO DIMENSIONAL

Para iniciar el modelado dimensional se debe tener en cuenta el primer objetivo de cualquier *Datamart*: el análisis de la información. Este análisis se realiza por medio de reportes, por lo tanto al modelar el *Datamart* se debe tener en cuenta la información que se desea ver en los reportes.

Dada la información de los requerimientos se identifican reportes académicos que tengan información de: inscripciones y admisiones, matrículas y cancelaciones, tanto de estudiantes como de materias, información de profesores. Se tienen en cuenta puntos importantes como: carreras, materias, periodos, ciudades a las que pertenecen los estudiantes, colegios de donde provienen los estudiantes, información personal de estudiantes y profesores, aprobación de materias, eventos sobre las materias (tales como cursos especiales, cursos de vacaciones, homologaciones, validaciones), eventos (tales como matrículas de honor, retiros, reingresos, entre otros), el estado del alumno (si es graduado, egresado o es no graduado), vigencia del alumno (si está matriculado o no), entre otras. Donde

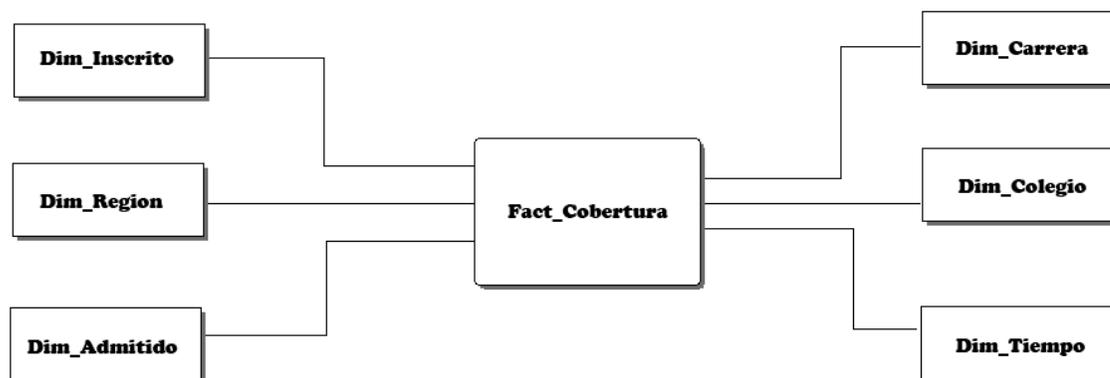
estas categorías, junto con las que se mencionará posteriormente son dimensiones en el modelo dimensional. Las tablas donde aparecen las medidas, como por ejemplo: cantidad de estudiantes admitidos, cantidad de estudiantes matriculados, cantidad de profesores, etc., son las tablas de hechos.

3.3.1 El Datamart. El *Datamart* diseñado en este proyecto tiene varias fuentes de datos: la base de datos “notas”, implementada en el SGBD *PostgreSQL*, la base de datos “profesores” también implementada en *PostgreSQL* y la base de datos “Registro, inscripciones y cartera” implementada en el SGBD *Informix*. Algunos detalles de estas fuentes de datos se encuentran en el DVD anexo, donde se describe el diccionario de datos de las bases de datos y los scripts de las fuentes utilizadas para el desarrollo del *Datamart*, el funcionamiento se describe en el anexo B.

El *Datamart* se diseñó adoptando el modelo conceptual de constelación definido en el párrafo 2.6.2. Se acogió este modelo debido a que el área académica maneja información en subáreas tales como se mencionó en el análisis de requerimientos, las cuales son área de inscripciones y admisiones, área de matrículas y población estudiantil, área de materias y rendimiento académico y por último el área de profesores, dando como resultado un análisis dimensional en múltiples tablas de hechos, definidas por estas diferentes categorías. Este esquema es una extensión del esquema en estrella, en el cuál las tablas de hechos comparten algunas tablas de dimensión.

En la figura 20 se define el modelo lógico del área de inscripciones y admisiones.

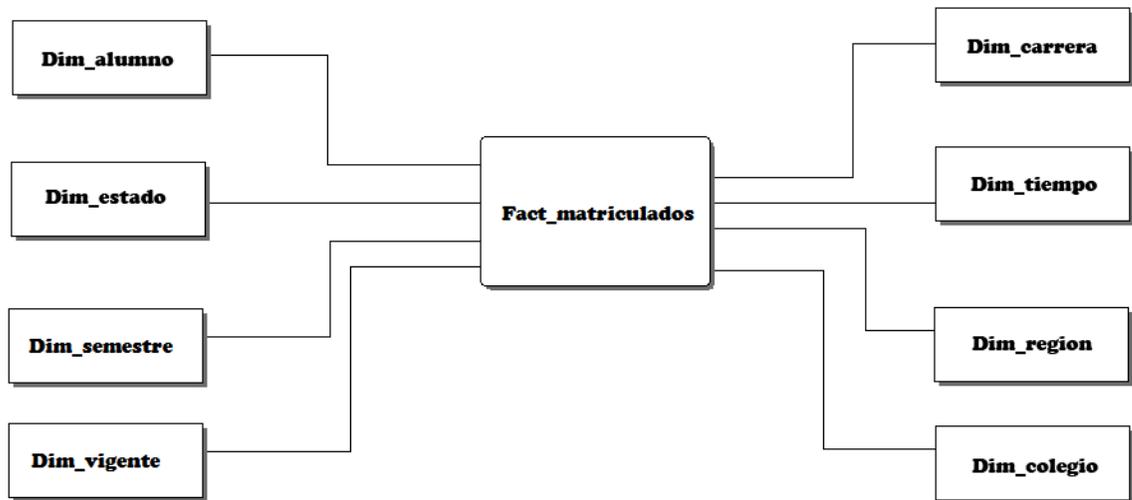
Figura 20. Cubo inscripciones y admisiones



Fuente. Esta investigación.

En la figura 21 se define el cubo de matrículas y población estudiantil, perteneciente al área de matrículas y población estudiantil.

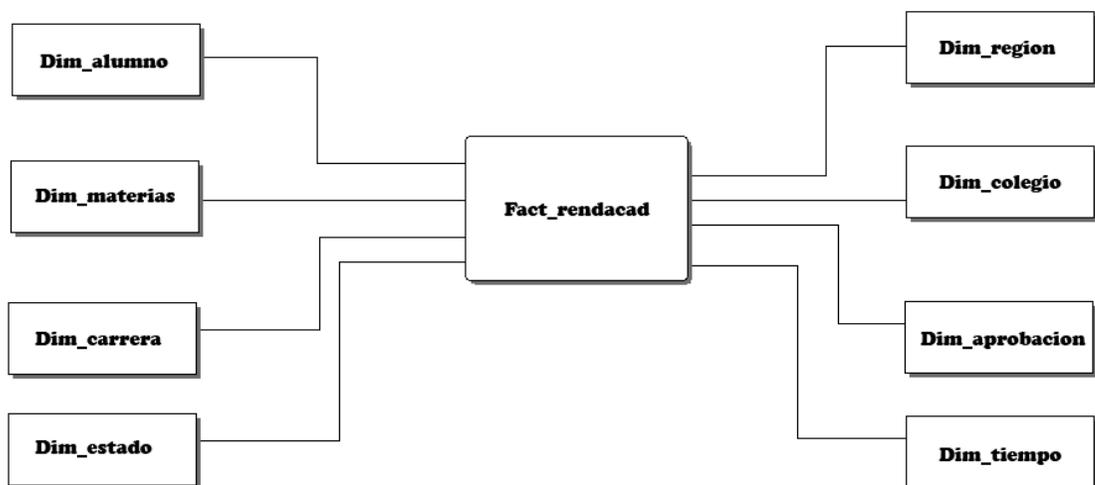
Figura 21. Cubo matrículas y población estudiantil



Fuente. Esta investigación.

En la figura 22 se define el cubo de rendimiento académico, perteneciente al área de materias y rendimiento académico.

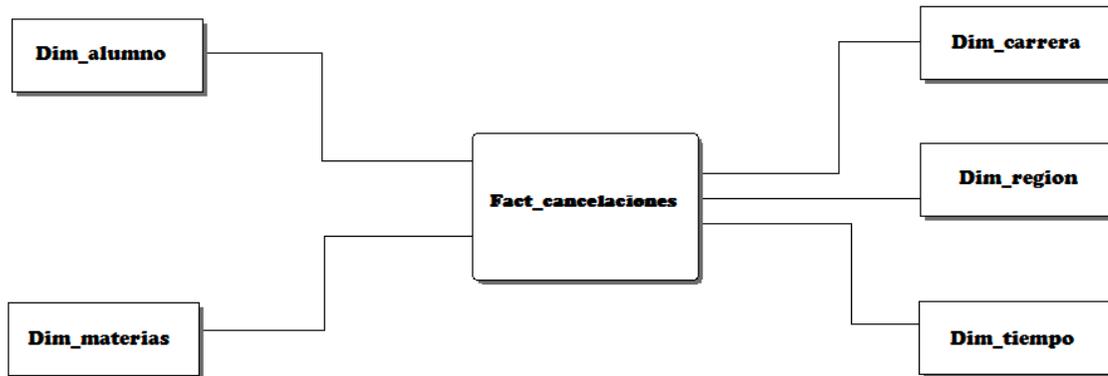
Figura 22. Cubo Materias y Rendimiento académico



Fuente. Esta investigación

En la figura 23 se define el cubo de cancelaciones, perteneciente al área de materias y rendimiento académico.

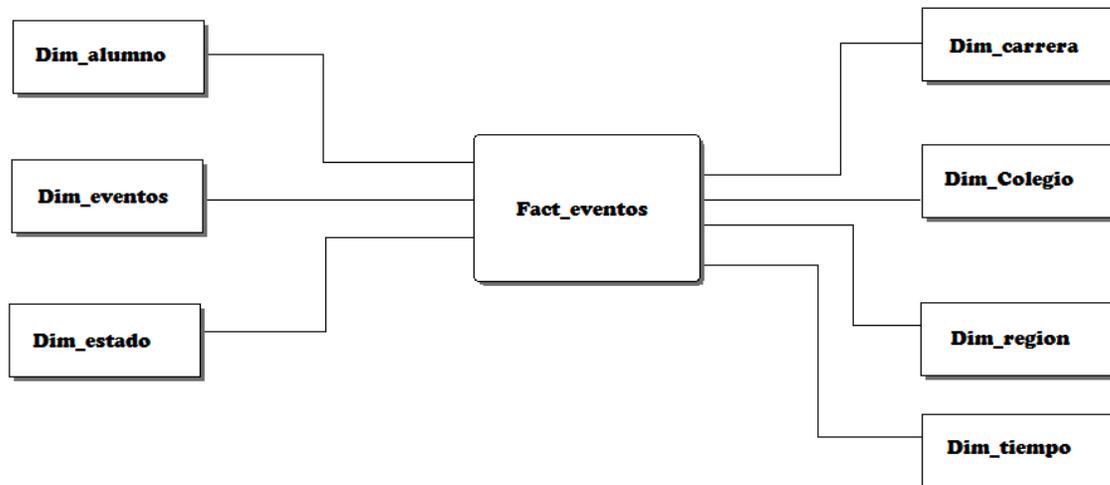
Figura 23. Cubo cancelaciones



Fuente. Esta investigación.

En la figura 24 se define el cubo de eventos de estudiantes, perteneciente al área de materias y rendimiento académico.

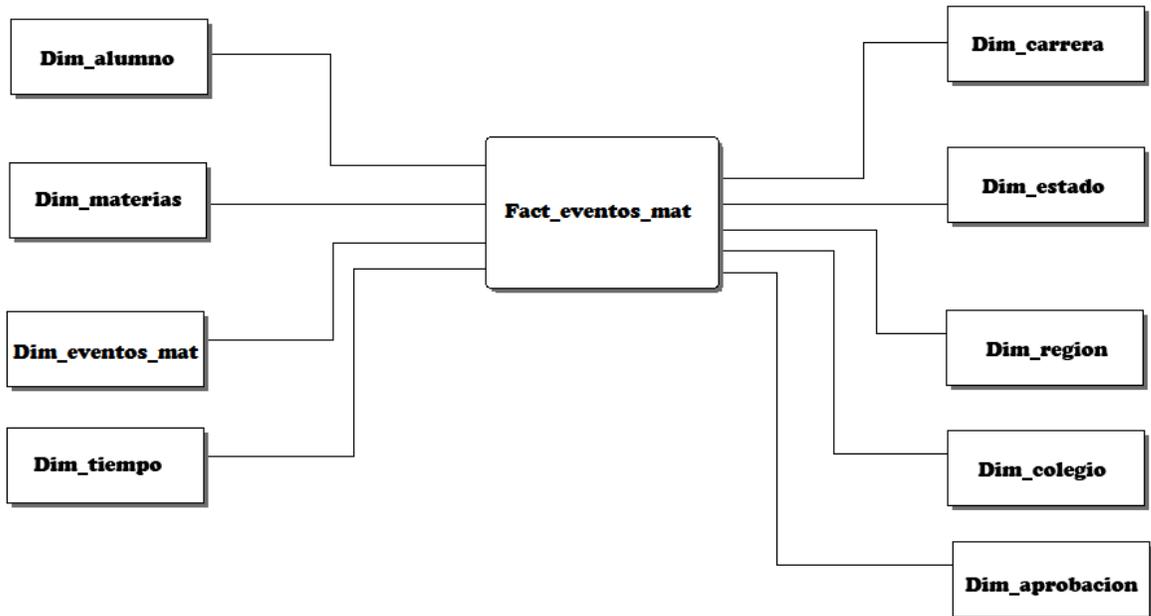
Figura 24. Cubo eventos estudiantes



Fuente. Esta investigación.

En la figura 25 se define el cubo de eventos de materias, perteneciente al área de materias y rendimiento académico.

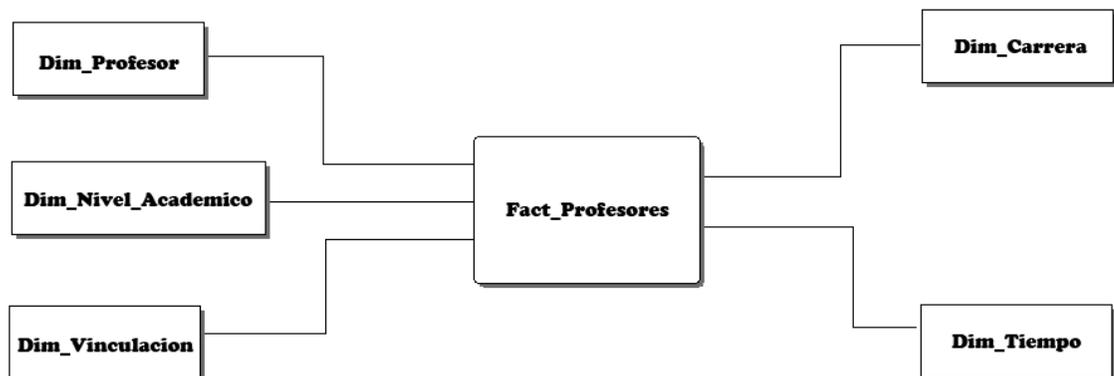
Figura 25. Cubo eventos materias



Fuente. Esta investigación.

En la figura 26 se define el cubo de profesores, perteneciente al área de profesores.

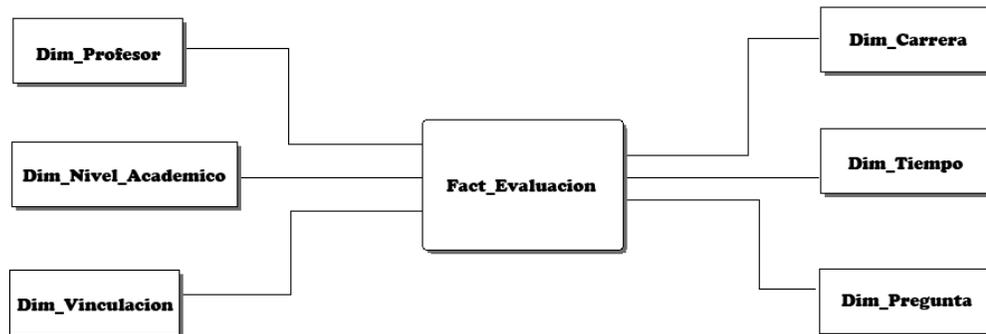
Figura 26. Cubo Profesores



Fuente. Esta investigación.

En la figura 27 se define el cubo de evaluaciones, perteneciente al área de profesores.

Figura 27. Cubo Evaluaciones



Fuente. Esta investigación.

3.3.2 Definición de granularidad. La granularidad está definida por las diferentes tablas de hechos de la siguiente forma:

3.3.2.1 Granularidad área de inscripciones y admisiones. La granularidad en la tabla de hechos “fact_cobertura”, responsable de esta área, corresponde a estudiantes inscritos o admitidos en un determinado semestre o periodo. Las medidas son número de inscritos, número de admitidos, ponderado de admisión y edad de los aspirantes, que son los campos de valor numérico, y sobre los cuales se hará operaciones de agregación sobre la tabla de hechos con la granularidad establecida.

3.3.2.2 Granularidad área de matrículas y población estudiantil. La granularidad en la tabla de hechos “fact_matriculados”, responsable del proceso de matrículas, corresponde al número de estudiantes en un determinado semestre o año. Como medida se identificó el número de alumnos sobre el cual se hará las operaciones de agregación, en este caso únicamente un conteo de los datos.

3.3.2.3 Granularidad área de matrículas de materias y rendimiento académico. En los siguientes puntos se definen los niveles de granularidad de las diferentes tablas de hechos diseñadas para soportar la información de materias y rendimiento académico.

- La granularidad en la tabla de hechos “fact_rendacad”, responsable de operaciones con notas y el número de materias matriculadas, corresponde a notas y/o materias matriculadas en un determinado semestre o año. Las medidas son número de materias matriculadas y notas, sobre los cuales se hará operaciones agregación sobre la tabla de hechos con la granularidad establecida.
- La granularidad en la tabla de hechos “fact_eventos_mat”, responsable de notas de eventos en materias tales como cursos especiales, cursos de verano, homologaciones o validaciones, corresponde a notas de estos eventos en un determinado semestre o año. Las medidas son el número de veces que se da el evento y la nota por cada evento, sobre los cuales se hará las operaciones de agregación sobre la tabla de hechos con la granularidad establecida.
- La granularidad en la tabla de hechos “fact_eventos”, responsable de eventos en estudiantes, como por ejemplo: matrículas de honor, pérdidas de semestre, reingresos, entre otros, corresponde a eventos dados en un determinado semestre o año. La medida es únicamente las veces que se da el evento, sobre la cual se hará las operaciones de agregación sobre la tabla de hechos con la granularidad establecida.
- La granularidad en la tabla de hechos “fact_cancelaciones”, responsable de cancelaciones de materias, corresponde a eventos dados en un determinado semestre o año. La medida es únicamente las veces que se cancela las materias, sobre la cual se hará las operaciones de agregación sobre la tabla de hechos con la granularidad establecida.

3.3.2.4 Granularidad Área de profesores. En los siguientes puntos se describe la granularidad de las tablas de hechos diseñadas para soportar la información de profesores y evaluación docente.

- La granularidad en la tabla de hechos “fact_profesores”, responsable de contar el número de profesores, corresponde a número de profesores en un determinado semestre o año. La medida es únicamente el número de profesores, sobre la cual se hará las operaciones de agregación sobre la tabla de hechos con la granularidad establecida.
- La granularidad de la tabla de hechos “fact_evaluacion”, responsable de las calificaciones en el proceso de evaluación docente, corresponde a las calificaciones en un determinado semestre o año. Las medidas son gran fortaleza,

fortaleza, transición, debilidad y gran debilidad, sobre las cuales se hará operaciones de agregación sobre la tabla de hechos con la granularidad establecida.

3.3.3 Dimensiones. A continuación se definen las dimensiones que soportan los requerimientos mencionados en la sección 3.2.

Las siguientes dimensiones son comunes entre las diferentes tablas de hechos diseñadas, posteriormente se definirá aquellas dimensiones que tratan un área específica.

3.3.3.1 Dimensiones comunes en el Datamart. Las dimensiones que se describen en este párrafo son dimensiones que se comparten entre las diferentes tablas de hechos diseñadas, para las áreas identificadas en el proceso de análisis de requerimientos.

- **Dimensión tiempo.** Esta dimensión contiene información al periodo de tiempo de los hechos. En la tabla 9 se muestra como está estructurada la dimensión tiempo. Esta dimensión es común para todas las áreas identificadas.

Tabla 9. Dimensión Tiempo

Dimensión	dim_tiempo	
Atributos	Tipo	Pk
id_tiempo	integer	Si
anio	integer	No
periodo	character varying(7)	No

Fuente. Esta investigación.

- **Dimensión alumno.** Esta dimensión contiene información relacionada con estudiantes de la Universidad de Nariño, esta se relaciona con las áreas de matrículas y población estudiantil, y materias e información de estudiantes. En la tabla 10 se muestra como está estructurada la dimensión alumno. Esta dimensión no se utiliza en el proceso de inscripciones y admisiones, ni en el proceso de profesores.

Tabla 10. Dimensión alumno

Dimensión		dim_alumno	
Atributos	Tipo	Pk	
id_alumno	integer	Si	
nombres	character varying(30)	Si	
apellidos	character varying(30)	No	
fecha_nacimiento	date	No	
Sexo	character (1)	No	
edad_ingreso	integer	No	
fecha_ingreso	date	No	
tipo_documento	character varying (50)		
estado_civil	character varying(20)	No	
tipo_sangre	character varying(12)	No	
estrato	character (2)	No	
cupo	character varying(50)	No	
victima_conflicto	character varying(11)	No	
etnia	character varying(50)	No	
capacidad	character varying(30)	No	
discapacidad	character varying(40)	No	

Fuente. Esta investigación.

- **Dimensión carrera.** Esta dimensión contiene información relacionada a los programas que ofrece la Universidad de Nariño, en la tabla 11 se muestra como está estructurada la dimensión carrera. Esta dimensión es común en todas las áreas.
- **Dimensión colegio.** Esta dimensión permite conocer información correspondiente a colegios a los cuales los aspirantes pertenecen en la tabla 12. se muestra como está estructurada la dimensión colegio. Esta dimensión no es común para el área de profesores.

Tabla 11. Dimensión carrera

Dimensión		dim_carrera
Atributos	Tipo	Pk
id_carrera	smallint	Si
facultad	character varying(50)	No
departamento	character varying(50)	No
carrera	character varying(90)	No
sede	character varying(10)	No
jornada	character varying(10)	No

Fuente. Esta investigación.

Tabla 12. Dimensión colegio

Dimensión		dim_colegio
Atributos	Tipo	Pk
id_colegio	integer	Si
departamento	character varying(30)	No
municipio	character varying(50)	No
nombre	character varying(100)	No
tipo	character varying(10)	No
calendario	character varying(10)	No
jornada	character varying(15)	No

Fuente: Esta investigación.

- **Dimensión región.** Esta dimensión permite conocer información correspondiente al lugar de procedencia de los aspirantes, admitidos y estudiantes, en la tabla 13 se muestra como se estructura la dimensión región.

- **Dimensión materias.** Esta dimensión contiene información de las materias en los diferentes programas, junto con el tipo de materia, tal como materias normales, electivas, competencias básicas y humanísticas. La tabla 14 muestra se estructura la dimensión. Esta dimensión no se encuentra presente en el área de inscripciones y admisiones, y matrículas y población estudiantil.

- **Dimensión estado.** Esta dimensión contiene información acerca alumnos tal como: alumno no graduado, egresado o graduado. Esta dimensión es común para el área de matrículas y población estudiantil y el área de materias y rendimiento académico. La tabla 15 describe la estructura de esta dimensión.

Tabla 13. Dimensión región

Dimensión	dim_region	
Atributos	Tipo	Pk
id_region	character varying(5)	Si
pais	character varying(35)	No
departamento	character varying(35)	No
municipio	character varying(35)	No

Fuente. Esta Investigación.

Tabla 14. Dimensión materias.

Dimensión	dim_materias	
Atributos	Tipo	Pk
id_materia	smallint	Si
detalle	character varying(75)	No
tipo_materia	character varying(35)	No

Fuente. Esta investigación.

Tabla 15. Dimensión estado

Dimensión	dim_estado	
Atributos	Tipo	Pk
id_estado	integer	Si
detalle_estado	character varying(20)	No

Fuente. Esta investigación.

3.3.3.2 Dimensiones área de inscripciones y admisiones. A continuación se describen las dimensiones diseñadas para el área de inscripciones y admisiones que junto con las dimensiones mencionadas anteriormente permitirán el análisis multidimensional.

- **Dimensión inscrito.** Esta dimensión contiene información personal de los aspirantes a las carreras en la Universidad de Nariño, en la tabla 16 se muestra como está estructurada la dimensión inscrito.

- **Dimensión admitido.** Esta dimensión sirve para determinar si el aspirante fue admitido o no. En la tabla 17 se muestra como está estructurada la dimensión admitido.

Tabla 16. Dimensión Inscrito

Dimensión		dim_inscrito
Atributos	Tipo	Pk
id_inscrito	character varying(14)	Si
tipo_documento	character varying(31)	No
nombres	character varying(30)	No
apellidos	character varying(30)	No
fecha_nacimiento	character varying(20)	No
sexo	character (1)	No
edad	smallint	No
estado_civil	character varying(20)	No
tipo_sangre	character varying(15)	No
estrato	character varying(20)	No
cupo	character varying(50)	No
validacion_icfes	character varying(11)	No
victima_conflicto	character varying(11)	No
etnia	character varying(50)	No
capacidad	character varying(30)	No
discapacidad	character varying(40)	No
tipo_ingreso	character varying(20)	No

Fuente: Esta investigación.

Tabla 17. Dimensión admitido

Dimensión		dim_admitido
Atributos	Tipo	Pk
id_admitido	smallint	Si
admitido	character varying(12)	No

Fuente: Esta investigación.

Las dimensiones: carrera, región, tiempo y colegio, mencionadas anteriormente en el párrafo 3.3.3.1, también hacen parte del área inscripciones y admisiones

3.3.3.3 Dimensiones de matrículas y población estudiantil. A continuación se describen cada una de las dimensiones que pertenecen al área de matrículas y población estudiantil.

- **Dimensión semestre.** Esta dimensión sirve para identificar el semestre en el que se encuentran los estudiantes. La tabla 18 muestra la estructura de esta dimensión.

- **Dimensión vigente.** Esta dimensión sirve para identificar si el estado de matrícula del alumno. La tabla 19 describe la estructura de esta dimensión.

Las dimensiones: alumno, estado, carrera, región, tiempo y colegio, mencionadas anteriormente en el párrafo 3.3.3.1, también hacen parte del área de matrículas y población estudiantil.

Tabla 18. Dimensión semestre

Dimensión		dim_semestre
Atributos	Tipo	Pk
id_semestre	integer	Si
Semestre	character varying(50)	No

Fuente. Esta investigación.

Tabla 19. Dimensión vigente

Dimensión	dim_vigente	
Atributos	Tipo	Pk
id_vigente	integer	Si
detalle_vigente	character varying(20)	No

Fuente. Esta investigación.

3.3.3.4 Dimensiones de materias y rendimiento académico. A continuación se describen cada una de las dimensiones que pertenecen al área de materias y rendimiento académico.

- **Dimensión aprobación.** Esta dimensión sirve para definir si los estudiantes aprobaron o no las materias, o eventos en las materias como cursos de vacaciones, cursos especiales, validaciones u homologaciones, entre otros. La tabla 20 muestra como está estructurada la dimensión aprobación.
- **Dimensión eventos.** Esta dimensión contiene los eventos que suceden sobre los estudiantes, como por ejemplo: matrículas de honor, retiros, entre otros. La tabla 21 muestra como está estructurada la dimensión eventos.
- **Dimensión eventos materias.** Esta dimensión almacena información de eventos sobre las materias tales como cursos especiales, cursos de verano, validaciones, homologaciones, entre otros. La tabla 22 muestra como está estructurada la dimensión de materias.

Tabla 20. Dimensión aprobación

Dimensión	dim_aprobacion	
Atributos	Tipo	Pk
id_aprobacion	integer	Si
detalle_aprobacion	character varying(15)	No

Fuente. Esta investigación.

Tabla 21. Dimensión eventos

Dimensión	dim_eventos	
Atributos	Tipo	Pk
id_evento	integer	Si
detalle_evento	character varying(75)	No

Fuente. Esta investigación.

Tabla 22. Dimensión eventos materias

Dimensión	dim_eventos_mat	
Atributos	Tipo	Pk
id_convencion	integer	Si
detalle_convencion	character varying (30)	No

Fuente. Esta investigación.

Las dimensiones: alumno, carrera, materias, colegio y región, mencionadas anteriormente en el párrafo 3.3.3.1 son también parte del área de materias y rendimiento académico.

3.3.3.5 Dimensiones área de profesores. A continuación se describen cada una de las dimensiones diseñadas para el área de profesores, en esta área se tiene en cuenta la información de profesores y evaluación docente.

- **Dimensión profesor.** Esta dimensión contiene la información personal de los profesores. En la tabla 23 se muestra como está estructurada la dimensión profesor.
- **Dimensión nivel académico.** Esta dimensión sirve para identificar el nivel académico del docente, tal como: doctor, magister, especialista o profesional. En la tabla 24 se muestra como está estructurada la dimensión nivel académico.

- **Dimensión Vinculación.** Esta dimensión contiene información de la vinculación de los docentes, como por ejemplo: si el docente es tiempo completo o es hora cátedra, entre otros. En la tabla 25 se muestra la estructura de esta dimensión.

Tabla 23. Dimensión profesor

Dimensión		dim_profesor	
Atributos	Tipo	Pk	
id_profesor	character varying(10)	Si	
nombres	character varying(30)	No	
apellidos	character varying(30)	No	
Sexo	char(1)	No	
Titulo	character varying(90)	No	

Fuente. Esta investigación.

Tabla 24. Dimensión nivel académico

Dimensión		dim_nivel_academico	
Atributos	Tipo	Pk	
id_nivel_academico	Smallint	Si	
nivel_academico	character varying(12)	No	

Fuente. Esta investigación.

Tabla 25. Dimensión vinculación

Dimensión		dim_vinculacion	
Atributos	Tipo	Pk	
id_vinculacion	smallint	Si	
vinculacion	character varying(40)	No	

Fuente. Esta investigación.

Las siguientes tablas pertenecen al proceso de evaluación docente:

- **Dimensión Pregunta:** Esta dimensión describe cada pregunta del proceso de evaluación docente. En la tabla 26 se muestra como está estructurada la dimensión pregunta.

Tabla 26. Dimensión pregunta

Dimensión		dim_pregunta	
Atributos	Tipo	Pk	
id_pregunta	smallint	Si	
formulario	character varying(120)	No	
caracteristica	character varying(70)	No	
variable	character varying(120)	No	

Fuente. Esta investigación.

La dimensión materias mencionada en el párrafo 3.3.3.1 también hace parte del área de profesores.

3.3.4 Tablas de hechos. El análisis multidimensional y la recolección de requisitos dio como resultado múltiples tablas hechos, clasificadas por las áreas: Inscripciones y admisiones, matrículas e información de estudiantes, matrículas de materias y rendimiento académico y por último el área de profesores. A continuación se clasifica las tablas de hechos por cada área.

3.3.4.1 Hechosinscripciones y admisiones. El análisis de inscripciones y admisiones dio como resultado la tabla “fact_cobertura”, descrita a continuación.

- **Hechoscobertura.** Esta tabla de hechos permite conocer información teniendo en cuenta atributos de medidas comocantidad de aspirantes, cantidad de admitidos, edad y puntaje de ingreso,agregados por año o periodo, programa, colegio y municipio.La descripción de esta tabla se muestra en la tabla 27.

3.3.4.2 Hechos matrículas y población estudiantil. Esta tabla de hechos permite conocer información de estudiantes teniendo en cuenta el número de alumnos, agregados por año o periodo, programa, semestre al que pertenecen,

identificando el estado de los estudiantes (no graduados, egresados o graduados), colegio y municipio. En la tabla 28 se muestra la descripción de esta tabla.

Tabla 27. Tabla de hechos cobertura

Nombre de la estructura del <i>Datamart</i>	Área de Inscripciones y admisiones		
Tabla de Hechos	fact_cobertura		
Atributos	Tipo	Pk	Descripción
id_inscrito	character varying(14)	Si	Identificador dimensión inscrito
id_carrera	smallint	Si	Identificador dimensión carrera
id_colegio	integer	Si	Identificador dimensión colegio
id_tiempo	smallint	Si	Identificador dimensión tiempo
id_region	character varying(5)	Si	Identificador dimensión región
id_admitido	smallint	Si	Identificador dimensión admitido
num_admitidos	integer	No	Número de admitidos
num_inscritos	integer	No	Número de inscritos
Edad	smallint	No	Edad inscrito
pond_total	numeric(7,3)	No	Ponderado total

Fuente. Esta investigación.

Tabla 28. Tabla de hechos matrículas y población estudiantil

Nombre de la estructura del <i>Datamart</i>	Área de Matrículas de matrículas y población estudiantil		
Tabla de Hechos	fact_matriculados		
Atributos	Tipo	Pk	Descripción
id_alumno	integer	Si	Identificador dimensión alumno
id_carrera	integer	Si	Identificador dimensión carrera
id_estado	integer	Si	Identificador dimensión estado
id_semestre	integer	Si	Identificador dimensión semestre
id_vigente	integer	Si	Identificador dimensión vigente
id_tiempo	integer	Si	Identificador dimensión tiempo
id_region	character(3)	Si	Identificador dimensión región
id_colegio	integer	Si	Identificador dimensión colegio

Tabla 28. Continuación

Atributos	Tipo	Pk	Descripción
num_alumnos	integer	No	Cantidad de alumnos

Fuente. Esta investigación.

3.3.4.3 Hechos matrículas de materias y rendimiento académico. El análisis de matrículas y rendimiento académico dio como resultado cuatro tablas de hechos, “fact_rendimiento”, “fact_eventos_mat”, “fact_eventos” y “fact_cancelaciones”. La estructura de cada una de estos hechos se define a continuación.

- **Tabla de hechos rendimiento académico y materias matriculadas.** Esta tabla de hechos permite conocer información de notas de estudiantes y materias matriculadas. La tabla 29 describe la estructura de esta tabla de hechos.

Tabla 29. Tabla de hechos rendimiento académico y materias matriculadas

Nombre de la estructura del <i>Datamart</i>		Área de Matrículas de materias y rendimiento académico	
Tabla de Hechos		fact_rendacad	
Atributos	Tipo	Pk	Descripción
id_alumno	integer	Si	Identificador dimensión alumno
id_materia	integer	Si	Identificador dimensión materia
id_carrera	integer	Si	Identificador dimensión carrera
id_estado	integer	Si	Identificador dimensión estado
id_region	character(5)	Si	Identificador dimensión región
id_colegio	Integer	Si	Identificador dimensión colegio
id_aprobacion	Integer	Si	Identificador dimensión aprobación
id_tiempo	Integer	Si	Identificador dimensión tiempo
Nota	double precision	No	Nota definitiva materia
num_matriculadas	Integer	No	Veces que se matricula una materia

Fuente. Esta investigación.

- **Tabla de hechos eventos sobre materias.** Esta tabla de hechos permite conocer información de eventos en las materias tales como: cursos especiales, cursos de vacaciones, homologaciones y validaciones. En la tabla 30 se describe la estructura de esta tabla de hechos.

Tabla 30. Tabla de hechos eventos sobre materias

Nombre de la estructura del <i>Datamart</i>		Área de Matrículas de materias y rendimiento académico	
Tabla de Hechos		fact_eventos_mat	
Atributos	Tipo	Pk	Descripción
id_alumno	Integer	Si	Identificador dimensión alumno
id_materia	Integer	Si	Identificador dimensión materia
id_carrera	Integer	Si	Identificador dimensión carrera
id_convencion	Integer	Si	Identificador dimensión eventos materias
id_estado	Integer	Si	Identificador dimensión estado
id_region	character(5)	Si	Identificador dimensión región
id_colegio	Integer	Si	Identificador dimensión colegio
id_aprobacion	Integer	Si	Identificador dimensión aprobación
id_tiempo	Integer	Si	Identificador dimensión tiempo
Nota	double precision	No	Nota definitiva materia
num_eventosmat	Integer	No	Veces que se registra un evento en las materias

Fuente. Esta investigación.

- **Tabla de hechos de eventos sobre estudiantes.** Esta tabla de hechos permite conocer información de eventos de estudiantes, como por ejemplo reingresos, matrículas de honor, retiros, etc. En la tabla 31 se describe la estructura de esta tabla.

- **Tabla de hechos cancelaciones.** Esta tabla de hechos permite conocer información de cancelaciones de materias. La estructura de esta tabla de hechos puede verse en la tabla 32.

Tabla 31. Tabla de hechos eventos sobre estudiantes

Nombre de la estructura del <i>Datamart</i>		Área de Matrículas de materias y rendimiento académico	
Tabla de Hechos		fact_eventos	
Atributos	Tipo	Pk	Descripción
id_alumno	integer	Si	Identificador dimensión alumno
id_evento	integer	Si	Identificador dimensión evento
id_carrera	integer	Si	Identificador dimensión carrera
id_estado	integer	Si	Identificador dimensión estado
id_region	character(5)	Si	Identificador dimensión región
id_colegio	integer	Si	Identificador dimensión colegio
id_tiempo	integer	Si	Identificador dimensión tiempo
num_eventos	integer	No	Veces que se da un evento

Fuente. Esta investigación.

Tabla 32. Tabla de hechos cancelaciones

Nombre de la estructura del <i>Datamart</i>		Área de Matrículas de materias y rendimiento académico	
Tabla de Hechos		fact_cancelaciones	
Atributos	Tipo	Pk	Descripción
id_alumno	Integer	Si	Identificador dimensión alumno
id_materia	Integer	Si	Identificador dimensión materia
id_carrera	Integer	Si	Identificador dimensión carrera
id_estado	Integer	Si	Identificador dimensión estado
id_region	character 5)	Si	Identificador dimensión región
id_tiempo	Integer	Si	Identificador dimensión tiempo
num_canceladas	Integer	No	Cantidad de materias canceladas

Fuente. Esta investigación.

3.3.4.4 Hechos profesores. El análisis de profesores dio como resultado dos tablas de hechos, “fact_profesores” y “fact_evaluacion”. La estructura y descripción de estos hechos se define a continuación.

Tabla 33. Tabla de hechos profesores

Nombre de la estructura del <i>Datamart</i>		Área de Profesores	
Tabla de Hechos		fact_profesores	
Atributos	Tipo	Pk	Descripción
id_profesor	character varying(10)	Si	Identificador dimensión inscrito
id_carrera	Smallint	Si	Identificador dimensión carrera
id_tiempo	Integer	Si	Identificador dimensión tiempo
id_nivel_academico	Smallint	Si	Identificador dimensión nivel académico
id_vinculacion	Smallint	Si	Identificador dimensión vinculación
num_profesores	Smallint	No	Número de profesores

Fuente. Esta investigación.

Tabla 34. Tabla de hechos evaluación docente

Nombre de la estructura del <i>Datamart</i>		Área de Evaluación	
Tabla de Hechos		fact_evaluacion	
Atributos	Tipo	Pk	Descripción
id_profesor	character varying(10)	Si	Identificador dimensión profesor
id_carrera	Smallint	Si	Identificador dimensión carrera
id_tiempo	Smallint	Si	Identificador dimensión tiempo
id_pregunta	Integer	Si	Identificador dimensión pregunta
id_nivel_academico	Smallint	Si	Identificador dimensión nivel académico
id_vinculacion	Smallint	Si	Identificador dimensión vinculación
gran_fortaleza	Smallint	No	cantidad Profesores con esta calificación

Tabla 34. Continuación

Atributos	Tipo	Pk	Descripción
Fortaleza	Smallint	No	cantidad Profesores con esta calificación
Transición	Smallint	No	cantidad Profesores con esta calificación
Debilidad	Smallint	No	cantidad Profesores con esta calificación
gran_debilidad	Smallint	No	cantidad Profesores con esta calificación

Fuente. Esta investigación.

3.3.5 Diseño del Modelo Dimensional. El modelo conceptual adoptado es el modelo en constelación, como se mencionó anteriormente, a continuación se describen los diferentes cubos de datos clasificados por cada área, los cuales conforman el *Datamart* de información académica.

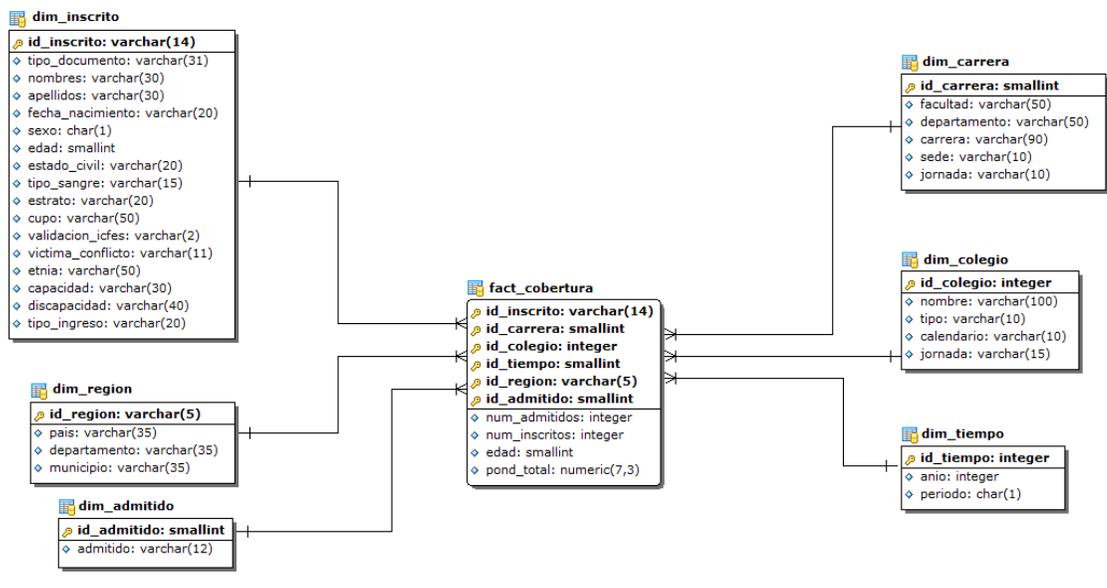
3.3.5.1 Modelo dimensional inscripciones y admisiones. En la figura 28 se muestra el cubo de datos que responde a las necesidades de inscritos y admitidos de forma detallada.

3.3.5.2 Modelado dimensional matrículas y población estudiantil. En la figura 29 se muestra el cubo de datos que responde a las necesidades de matrículas y población estudiantil de forma detallada.

3.3.5.3 Modelado dimensional materias y rendimiento académico. En la figura 30 se muestra el cubo que responde a las necesidades de materias y rendimiento académico.

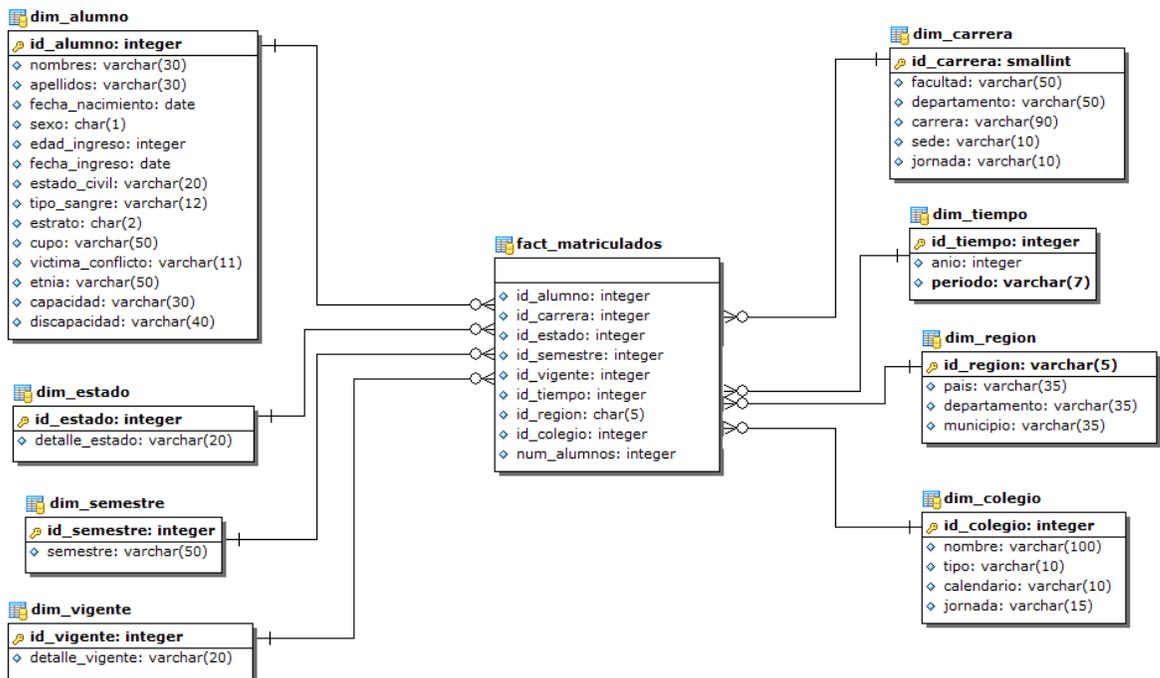
En la figura 31 se muestra el cubo que responde a las necesidades de materias canceladas.

Figura 28. Cubo inscripciones y admisiones detallado



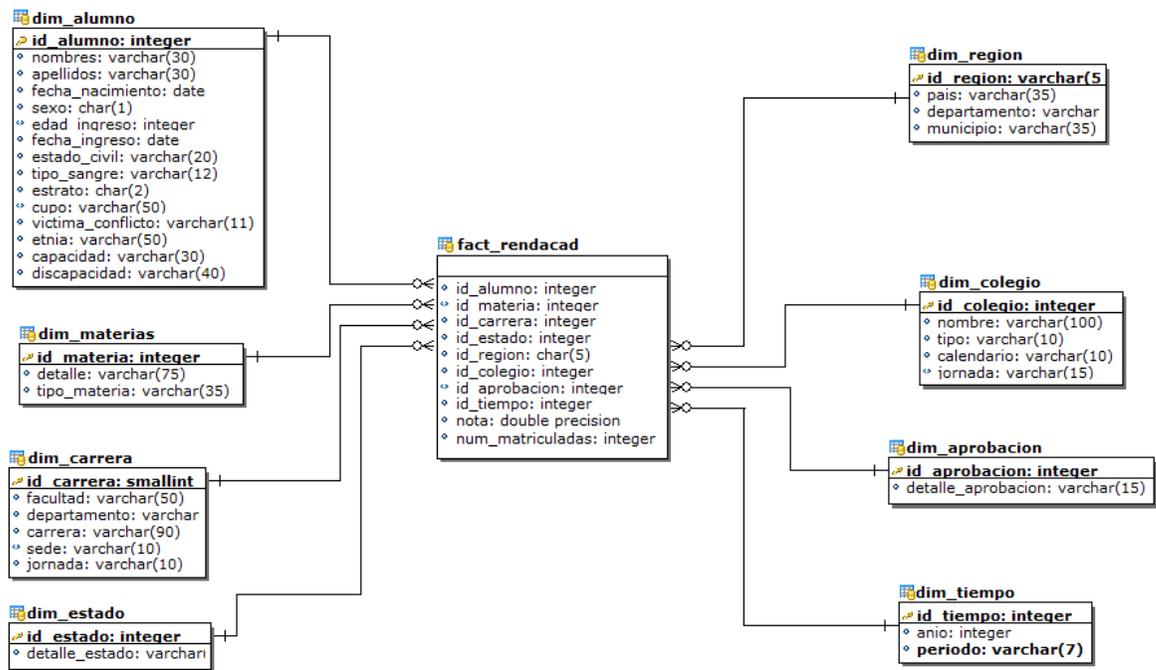
Fuente. Esta investigación.

Figura 29. Cubo de matrículas y población estudiantil detallado



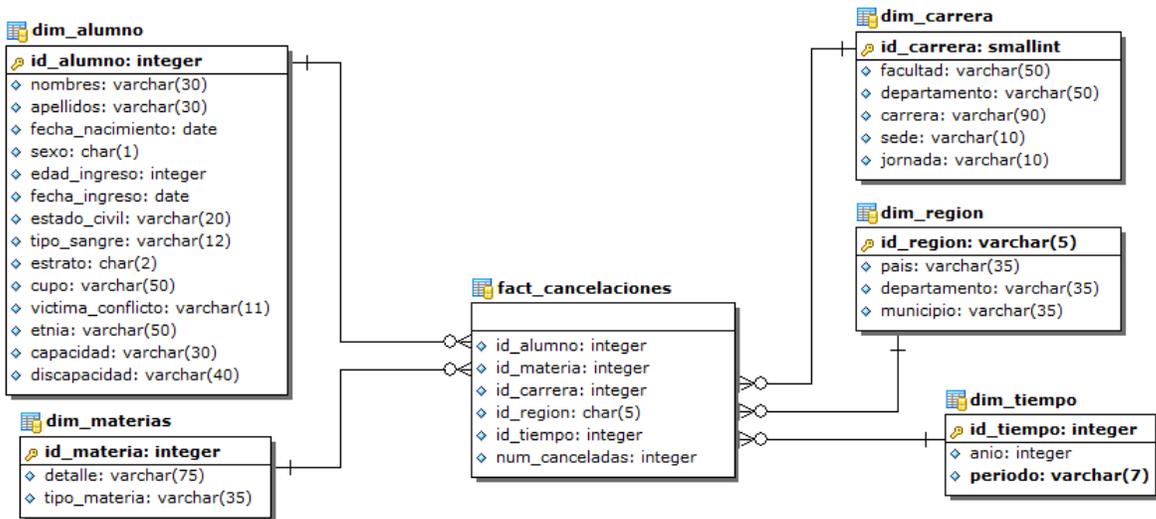
Fuente. Esta investigación.

Figura 30. Cubo materias y rendimiento académico detallado.



Fuente. Esta investigación.

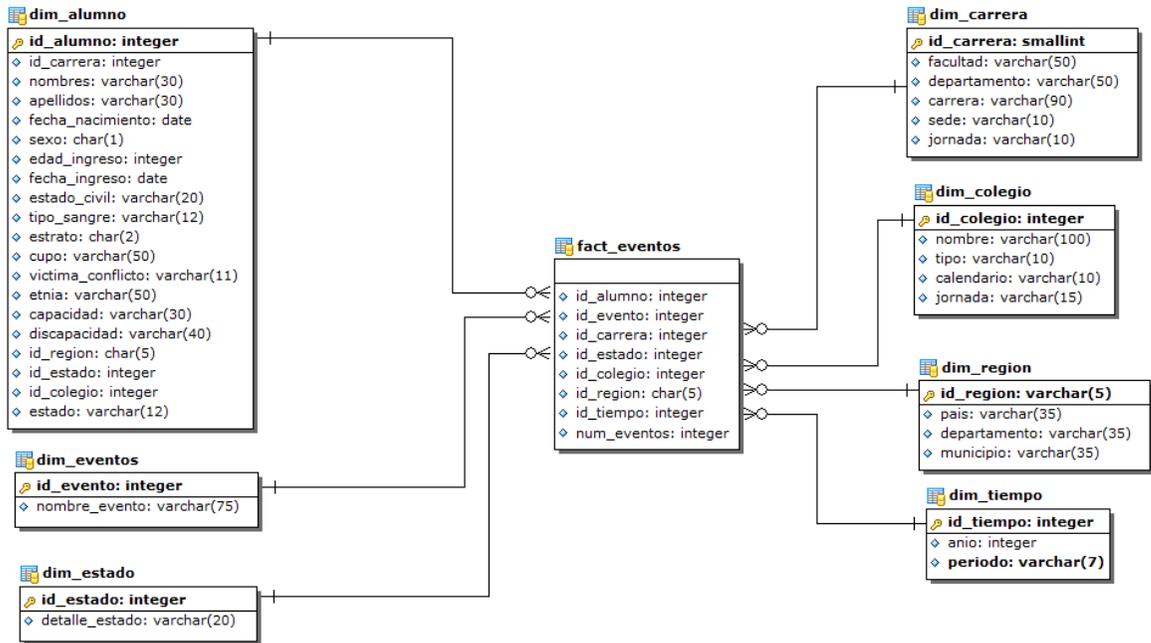
Figura 31. Cubo materias canceladas detallado



Fuente. Esta investigación

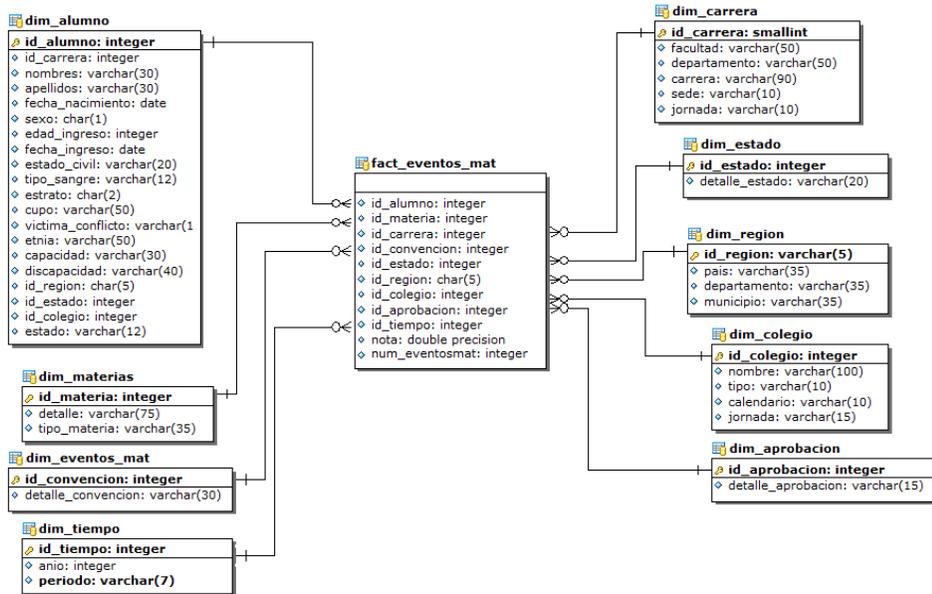
En la figura 32 se muestra el cubo que responde a las necesidades de eventos efectuados por los estudiantes.

Figura 32. Cubo eventos estudiantes detallado



Fuente. Esta investigación.

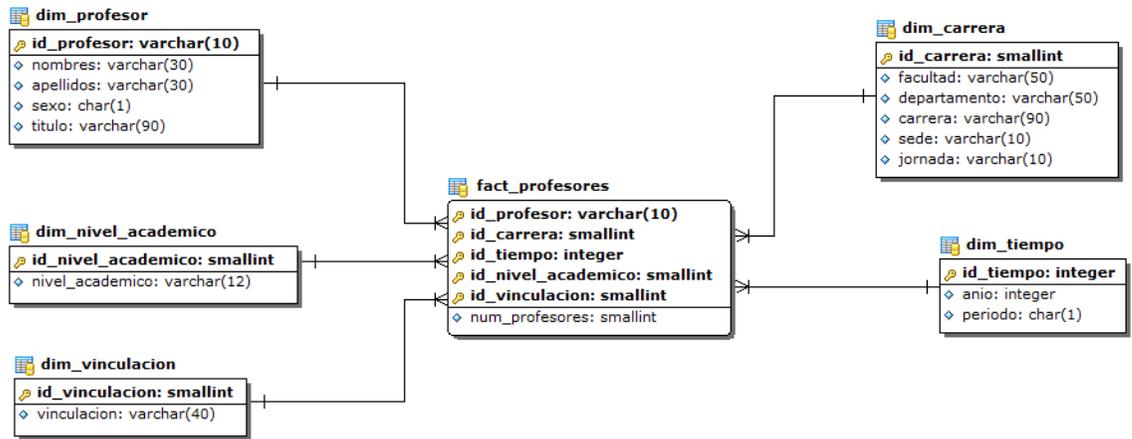
Figura 33. Cubo eventos materias detallado



Fuente. Esta investigación.

3.3.5.4 Modelado dimensional profesores. En la figura 34 se muestra el cubo que responde a las necesidades de información de profesores.

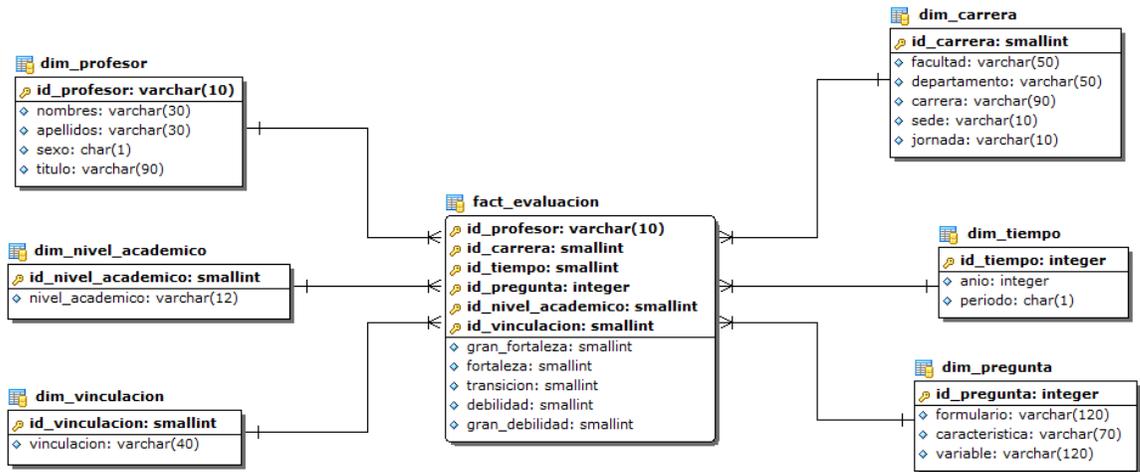
Figura 34. Cubo de profesores detallado



Fuente. Esta investigación.

En la figura 35 se muestra el cubo que responde a las necesidades de evaluaciones.

Figura 35. Cubo de evaluaciones detallado



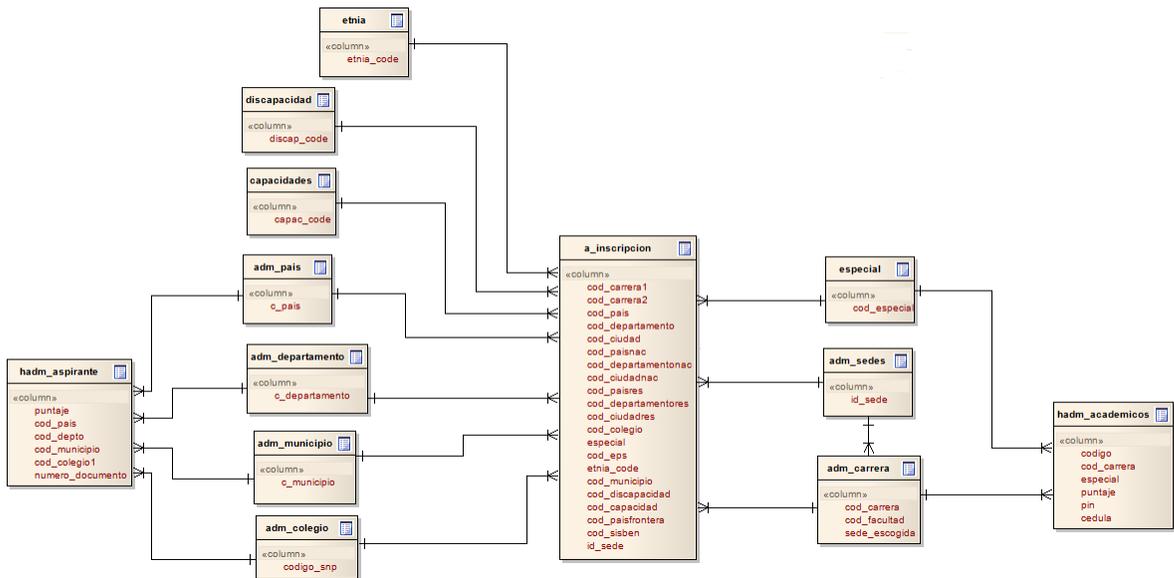
Fuente. Esta investigación.

3.4 DISEÑO TÉCNICO DE LA ARQUITECTURA

3.4.1 Datos. Para el análisis de los datos, se comienza por estudiar los datos fuentes que se manejan dentro de los procesos académicos de la Universidad de Nariño, el gestor en el que están implementadas las bases de datos, en este caso el Sistema Gestor de Bases de Datos - SGBD *PostgreSQL* y el Sistema Gestor de Bases de Datos- SGBD *Informix*, y a su vez se analiza la estructura de las tablas contenidas en las bases de datos de estos gestores. La Universidad de Nariño tiene bases de datos relacionales, las cuales se pueden expresar en diagramas Entidad-Relación en sus sistemas OLTP. Debido a la gran complejidad de este sistema de bases de datos se ha dividido el modelado de las bases de datos en diferentes diagramas de Entidad-Relación identificados por los diferentes procesos analizados anteriormente.

En la figura 36 se muestra el diagrama Entidad-Relación, E-R del área de inscripciones y admisiones de la Universidad de Nariño.

Figura 36. Modelo E-R área de inscripciones y admisiones

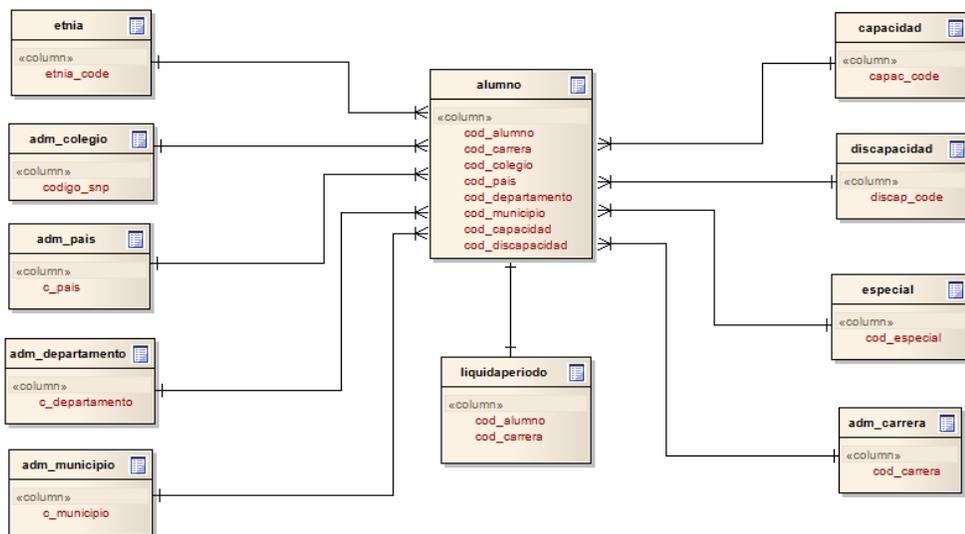


Fuente. Esta investigación.

En la figura 37 se muestra el diagrama E-R del área de matrículas y población estudiantil de la Universidad de Nariño.

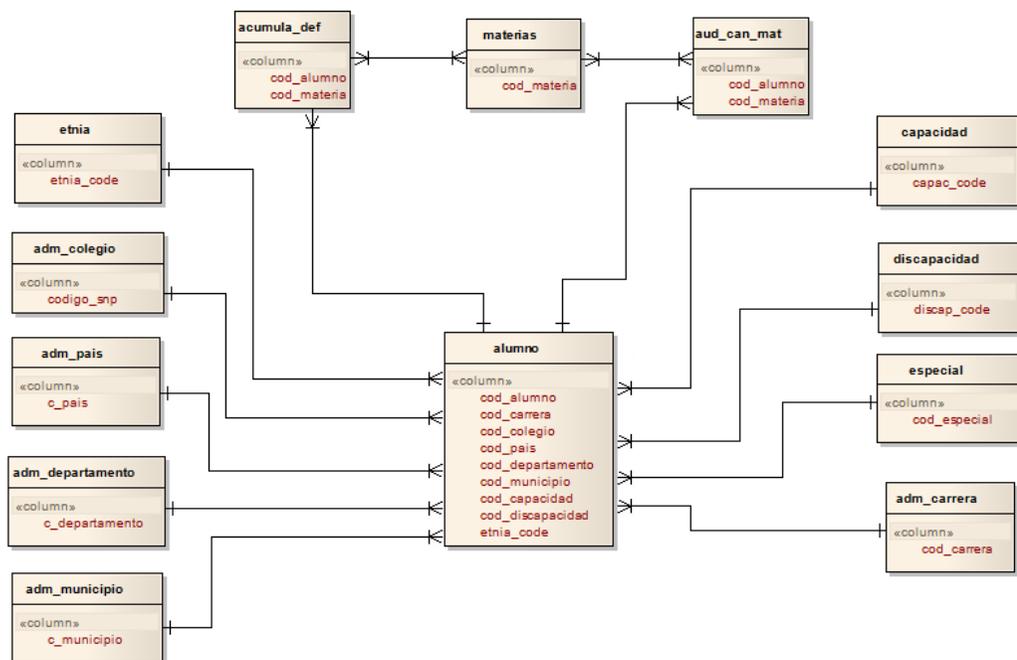
En la figura 38 se muestra el diagrama E-R del área de materias y rendimiento académico de la Universidad de Nariño.

Figura 37. Modelo E-R área de matrículas y población estudiantil.



Fuente. Esta investigación.

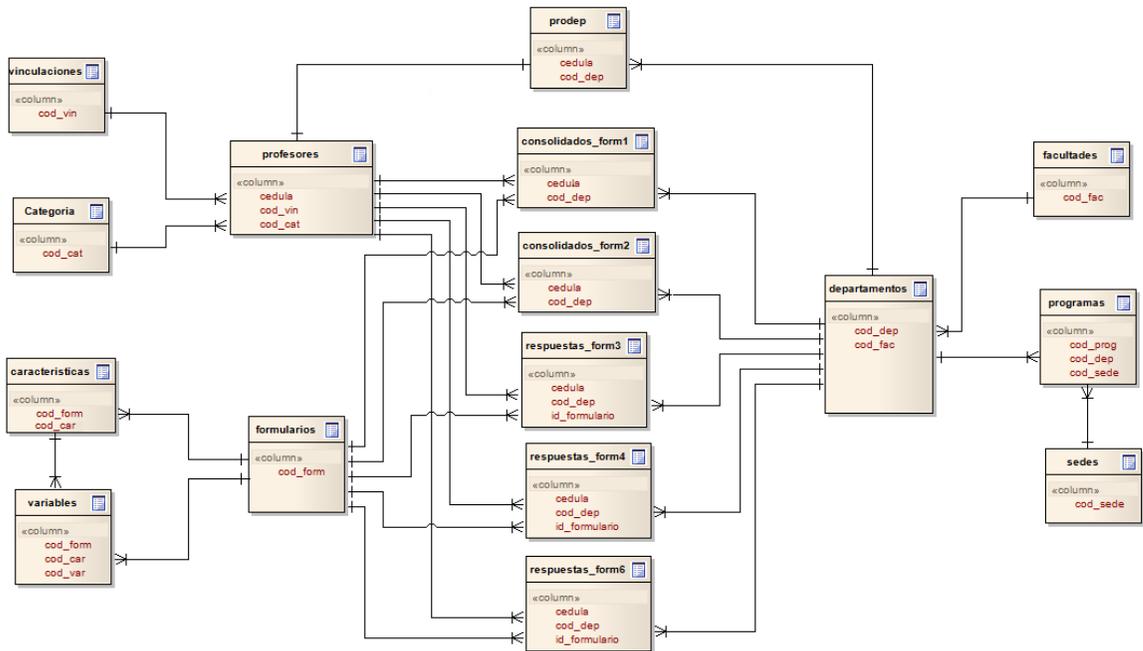
Figura 38. Modelo E-R área de materias y rendimiento académico



Fuente. Esta investigación.

En la figura 39 se muestra el diagrama E-R del área de profesores de la Universidad de Nariño.

Figura 39. Modelo E-R área de profesores.



Fuente. Esta investigación.

Para cargar los datos en el modelo dimensional se requiere la información de las tablas del sistema OLTP mencionadas anteriormente, en esta sección se describe en cada una de las dimensiones diseñadas, la respectiva fuente de datos y el tipo de dato que debe almacenar, además operaciones soportadas por cada atributo de las tablas de hechos que conforman el *Datamart*.

3.4.1.1 Mapeo dimensiones comunes. A continuación se muestra una descripción de los atributos y las fuentes de las dimensiones que comparten las tablas de hechos de las áreas identificadas anteriormente.

- **Dimensión Tiempo.** En la tabla 35 se detalla de una forma específica los datos fuente utilizados para creación de la dimensión tiempo además de una breve descripción de cada atributo que la conforma.

Los datos de la dimensión tiempo se deben mapear de la siguiente forma:

Tabla 35. Mapeo Dimensión tiempo

Nombre de la estructura del <i>Datamart</i>		Dimensiones comunes
Dimensión		dim_tiempo
Atributos	Fuente	Descripción
id_tiempo	Esta Investigación	Identificador del Semestre
anio	Esta Investigación	Año
periodo	Esta Investigación	Periodo Académico

Fuente. Esta investigación.

- **id_tiempo.** El atributo almacenará el código de identificación del periodo.
 - **anio.** El atributo almacenará el año.
 - **periodo.** El atributo almacenará el periodo, los datos aceptados para el atributo son: A, B,- ANUAL-.
- **Dimensión alumno.** En la tabla 36 se detalla de una forma específica los datos fuente utilizados para la creación de la dimensión alumno, además de una breve descripción de cada atributo que la conforma.

Tabla 36. Mapeo dimensión alumno

Nombre de la estructura del <i>Datamart</i>		Dimensiones comunes
Dimensión		dim_alumno
Atributos	Fuente	Descripción
id_alumno	alumno.cod_alumno	Identificador o código del alumno
nombres	alumno.nombres	Nombres del Estudiante
apellidos	alumno.apellidos	Apellidos del Estudiante
fecha_nacimiento	alumno.fecha_nac	Fecha de nacimiento
sexo	alumno.sexo	Sexo del estudiante
edad_ingreso	Esta investigación	Edad de ingreso del estudiante
fecha_ingreso	alumno.fecha_ingreso	Fecha de Ingreso del estudiante

Tabla 36. Continuación

Atributos	Fuente	Descripción
tipo_documento	alumno.tipo_documento	Tipo de documento de identificación
estado_civil	alumno.estado_civil	Estado Civil del estudiante
tipo_sangre	alumno.tipo_sangre	Tipo de Sangre del Estudiante
estrato	alumno.estrato	Estrato del Aspirante
cupo	especial.detalle_especial1	Cupo del Estudiante
victima_conflicto	a_inscripcion.victima_conflicto	Estudiante es Víctima de Conflicto
etnia	etnia.etnia_descr	Etnia que pertenece Estudiante
capacidad	capacidades.capac_descr	Capacidad del Estudiante
discapacidad	discapacidad.discap_descr	Discapacidad del Estudiante

Fuente. Esta investigación.

- **id_alumno.** El atributo almacenará el código o el identificador del alumno.
- **nombres.** El atributo almacenará los nombres de los estudiantes.
- **apellidos.** El atributo almacenará los apellidos de los estudiantes.
- **fecha_nacimiento.** El atributo almacenará la fecha de nacimiento de cada uno de los estudiantes
- **sexo.** El atributo almacenará el género del aspirante, los datos aceptados son: M, para sexo masculino, F, para sexo femenino.
- **edad_ingreso.** El atributo almacenará la edad de ingreso de los estudiantes. La edad de ingreso se calcula a partir de la fecha de nacimiento y la fecha de ingreso de los estudiantes.
- **fecha_ingreso.** El atributo almacenará la fecha de ingreso de los estudiantes.
- **tipo_documento.** El atributo almacenará el tipo de documento del aspirante, los datos aceptados para el atributo son: CEDULA DE CIUDADANIA, TARJETA DE IDENTIDAD, PASAPORTE, CEDULA EXTRANJERA, NUMERO UNICO DE IDENTIFICACION, para los aspirantes que no se conozca el tipo de documento se deberá ingresar el valor DESCONOCIDO.
- **estado_civil.** El atributo almacenará el estado civil del aspirante, los datos aceptados para el atributo son: CASADO, MADRE SOLTERA, RELIGIOSO, SEPARADO, SOLTERO UNION LIBRE, VIUDO. Para los aspirantes que no se conozca su estado civil se deberá ingresar el valor DESCONOCIDO.

- **tipo_sangre.** El atributo almacenará el tipo de sangre al que pertenece el aspirante, los datos aceptados para el atributo son: A+, A-, AB+, AB-, B+, B-, O+, O-. Para los aspirantes que no se conozca el tipo de sangre se deberá ingresar el valor DESCONOCIDO.
- **estrato.** El atributo almacenará el estrato el cual pertenece el aspirante, los datos aceptado para el atributo son de 0 a 10. Para los aspirantes que no se conozca su estrato se deberá ingresar el valor DESCONOCIDO.
- **cupo.** El atributo almacenará el tipo de cupo por el cual el aspirante ingresa, los datos aceptados para el atributo son: CUPO REGULAR, DEPORTISTA DESTACADO, DESPLAZADOS NARIÑO-PUTUMAYO, DISCAPACITADOS, EXTRANJERO SIN NACIONALIDAD COLOMBIANA, HIJO DE VICTIMA DE SECUESTRO, HIJO DE EMPLEADO, HIJO DE VICTIMA DE DESAPARICION, MEJOR BACHILLER DEPARTAMENT DE NARIÑO, MUNICIPIOS DEPRIMIDOS SOCIO-ECONOMICAMENTE, NEGRITUDES DE LA ZONA PACIFICA NARIÑENSE, PERTENECIENTES A CABILDOS INDIGENAS DE NARIÑO, PERTENECIENTE CABILDO INDIGENA PUTUMAYO, PROFESIONALES, REINSERTADOS.
- **victima_conflicto.** El atributo almacenará información que permita conocer si el aspirante es víctima de conflicto, los datos aceptados para el atributo son SI, NO. Para los aspirantes que no se conozca esta información se deberá ingresar el valor DESCONOCIDO.
- **etnia.** El atributo almacenará la etnia que pertenece el aspirante, los datos aceptados para el atributo son: ACHAGUA, AWA(CUAIKER), COFÁN O KOFÁN, CHIRICOA, EMBERA CAITO O KATÍO, EMBERA CHAMI, EMPERARA SIAPIDARA, GUAMBIANO, IGNA, KAMSA O KAMENTSÁ, NASA (PAÉZ), NEGRITUDES, PASTOS, PIAPOCO (DZASE), PIAROA, PIRATAPUTO, QUILLACINGAS, WITOTO, YANACONA, YUKUNA, ZENÚ. Para los aspirantes que no pertenezcan a ninguna etnia se deberá ingresar el valor NO APLICA, además a los aspirantes que no se conozcan su etnia está el valor DESCONOCIDO.
- **capacidad.** El atributo almacenará la capacidad del aspirante, los datos aceptados para el atributo son: CON TALENTO CIENTIFICO, CON TALENTO SUBJETIVO, CON TALENTO TECNOLOGICO, SUPERDOTADO. Para los aspirantes que no posean ninguna capacidad se deberá ingresar el valor NO APLICA, además para los aspirantes que no se conozca esta información se encuentra el valor DESCONOCIDO.
- **discapacidad.** El atributo almacenará la discapacidad del aspirante, los datos aceptados para el atributo son: AUTISMO, BAJA VISION DIAGNOSTICA, CEGUERA, DEFICIENCIA COGNITIVA (RETARDO

MENTAL), HIPOACUSIA A BAJA AUDICION, LESION NEUROMUSCULAR, MULTIPLE, PARALISIS CEREBRAL, SINDROME DE DOWN, SORDERA PROFUNDA. Para los aspirantes que no posean ninguna discapacidad se deberá ingresar el valor NO APLICA, además para los aspirantes que no se conozca esta información se encuentra el valor DESCONOCIDO.

- **Dim_carrera.** En la tabla 37 se detalla de una forma especificada los datos fuente utilizados para creación de la dimensión carrera además de una breve descripción de cada atributo que la conforma.

Tabla 37. Mapeo dimensión carrera

Nombre de la estructura del <i>Datamart</i>		Dimensiones comunes
Dimensión		dim_carrera
Atributos	Fuente	Descripción
id_carrera	programas.cod_prog	Identificador Carrera
facultad	facultades.nom_fac	Nombre de Facultad
departamento	departamentos.nom_dep	Nombre del Departamento
carrera	programas.nom_prog	Nombre de la Carrera
sede	sedes.sede	Sede de la Carrera
jornada	programas.jornada	Jornada de la Carrera

Fuente. Esta investigación.

Los datos de la dimensión carrera se deben mapear de la siguiente forma:

- **id_carrera.** El atributo almacenará el código de identificación de la carrera.
- **facultad.** El atributo almacenará el nombre de la facultad que pertenece la carrera, el valor debe estar en mayúscula.
- **departamento.** El atributo almacenará el nombre del departamento que pertenece la carrera, el valor debe estar en mayúscula.
- **carrera.** El atributo almacenará el nombre de la carrera, el valor debe estar en mayúscula.
- **sede.** El atributo almacenará el nombre de la sede donde se ofrece la carrera, el valor debe estar en mayúscula.

- **jornada.** El atributo almacenará información que permita conocer la jornada en la cual se ofrece la carrera, los datos aceptados para el atributo son: DIURNA, NOCTURNA.
- **Dimensión colegio.** En la tabla 38 se detalla de una forma específica los datos fuente utilizados para creación de la dimensión colegio además de una breve descripción de cada atributo que la conforma.

Tabla 38. Dimensión colegio

Nombre de la estructura del <i>Datamart</i>		Dimensiones comunes
Dimensión		dim_colegio
Atributos	Fuente	Descripción
id_colegio	adm_colegio.codigosnp	Identificador Colegio
departamento	adm_colegio.departamento	Nombre Departamento
municipio	adm_colegio.municipio	Nombre Municipio
nombre	adm_colegio.colegio	Nombre Colegio
tipo	adm_colegio.tipo_colegio	Tipo de Colegio
calendario	adm_colegio.calendario	Calendario del Colegio
jornada	adm_colegio.jornada	Jornada del Colegio

Fuente. Esta investigación.

Los datos de la dimensión colegio se deben mapear de la siguiente forma:

- **id_colegio.** El atributo almacenará el código de identificación del colegio.
- **departamento.** El atributo almacenará el departamento al que pertenece el colegio.
- **municipio.** El atributo almacenará el municipio al que pertenece el colegio
- **nombre.** El atributo almacenará el nombre del colegio de donde proviene el aspirante y/o estudiante, el valor debe estar en mayúscula.
- **tipo.** El atributo almacenará el tipo de colegio, los datos aceptados para el atributo son: OFICIAL, NO OFICIAL.
- **calendario.** El atributo almacenará el calendario del colegio, los datos aceptados para el atributo son: A, B, FLEXIBLE.

- **jornada.** El atributo almacenará la jornada del colegio, los datos aceptados para el atributo son: COMPLETA, DIURNA, NOCTURNA, SABATINA, TARDE.

• **Dimensión región.** En la tabla 39 se detalla de una forma específica los datos fuente utilizados para creación de la dimensión región además de una breve descripción de cada atributo que la conforma.

Tabla 39. Mapeo dimensión región

Nombre de la estructura del <i>Datamart</i>		Dimensiones comunes
Dimensión		dim_region
Atributos	Fuente	Descripción
id_region	adm_municipio.c_codigonacional	Identificador Municipio
pais	adm_pais.c_detalle_pais	Nombre País
departamento	adm_departamento.c_nombre_depto	Nombre Departamento
municipio	adm_municipio.c_detalle_mcpio	Nombre Municipio

Fuente. Esta investigación.

Tabla 40. Mapeo dimensión materias

Nombre de la estructura del <i>Datamart</i>		Dimensiones comunes
Dimensión		dim_materias
Atributos	Fuente	Descripción
id_materia	materias.cod_materia	Identificador o Código de Materia
detalle	materias.detalla_largo	Descripción de materias
tipo_materia	Esta Investigación	Tipo de Materia

Fuente. Esta investigación.

- **id_region.** El atributo almacenará el código de identificación del municipio.
- **pais.** El atributo almacenará el nombre del país del aspirante y/o estudiante, el valor debe estar en mayúscula.

- **departamento.** El atributo almacenará el nombre del departamento del aspirante y/o estudiante, el valor debe estar en mayúscula.
- **municipio.** El atributo almacenará el nombre del municipio del aspirante y/o estudiante, el valor debe estar en mayúscula.

• **Dimensión Materias.** En la tabla 40 se detalla de una forma específica los datos fuente utilizados para la creación de la dimensión materias, además de una breve descripción de cada atributo que la conforma.

Los datos de la dimensión materia se deben mapear de la siguiente forma:

- **id_materia.** El atributo almacenará el código de la materia.
- **detalle.** El atributo almacenará el nombre de la materia.
- **tipo_materia.** El atributo almacenará el tipo de materia, los datos aceptados para el atributo son: NORMAL, ELECTIVA, COMPETENCIA BÁSICA, HUMANÍSTICA.

• **Dimensión estado.** En la tabla 41 se detalla de una forma específica los datos fuente utilizados para la creación de la dimensión estado, además de una breve descripción de cada atributo que lo conforma.

Tabla 41. Mapeo dimensión estado

Nombre de la estructura del <i>Datamart</i>	Área de Matrículas de materias y rendimiento académico	
Dimensión	dim_estado	
Atributos	Fuente	Descripción
id_estado	Esta investigación	Identificador estado del estudiante
detalle_estado	Esta investigación	Descripción estado estudiante

Fuente. Esta investigación.

Los datos de la dimensión estado se deben mapear de la siguiente forma:

- **id_estado.** El atributo almacenará el identificador del estado del alumno.
- **detalle_estado.** El atributo almacenará el detalle del estado del estudiante tal como: NO GRADUADO, EGRESADO Y NO GRADUADO.

3.4.1.2 Mapeo inscripciones y admisiones. Para esta área se describe la tabla de hechos “fact_cobertura” y las dimensiones relacionadas, de acuerdo a los siguientes puntos:

- **Tabla de hechos cobertura.** En la tabla 42 se describe las posibles operaciones que se pueden realizar sobre cada una de las métricas o medidas de la tabla de hechos cobertura, los campos vacíos corresponden las operaciones que no se pueden realizar sobre la tabla (esto es para todas las tablas de hechos que se describen de aquí en adelante).

Tabla 42. Operaciones de la tabla de hechos cobertura

Tabla de hechos	fact_cobertura				
Atributo 1	num_admitidos				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_inscrito	Si				
dim_carrera	Si				
dim_colegio	Si				
dim_tiempo	Si				
dim_region	Si				
dim_admitido	Si				
Atributo 2	num_inscritos				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_inscrito	Si	Si			
dim_carrera	Si	Si			
dim_colegio	Si	Si			
dim_tiempo	Si	Si			
dim_region	Si	Si			
dim_admitido	Si	Si			

Tabla 42. Continuación

Atributo 3		edad			
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_inscrito			Si	Si	Si
dim_carrera			Si	Si	Si
dim_colegio			Si	Si	Si
dim_tiempo			Si	Si	Si
dim_region			Si	Si	Si
dim_admitido			Si	Si	Si

Atributo 4		pond_total			
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_inscrito			Si	Si	Si
dim_carrera			Si	Si	Si
dim_colegio			Si	Si	Si
dim_tiempo			Si	Si	Si
dim_region			Si	Si	Si
dim_admitido			Si	Si	Si

Fuente. Esta investigación.

Las medidas en la tabla de hechos cobertura deben ser valores numéricos, los datos correctos deben mapearse de la siguiente forma:

- **num_admitidos.** La métrica almacenará el número de aspirantes admitidos, los datos aceptados para el atributo son 0 y 1.
- **num_inscritos.** La métrica almacenará el número de aspirantes, el único valor aceptado para el atributo es 1.
- **edad.** La métrica almacenará la edad de los aspirantes por ejemplo: 19.
- **pond_total.** La métrica almacenará ponderado que obtiene el aspirante para ingresar a un determinado programa por ejemplo: 51.25.

• **Dimensión inscripción.** En la tabla 43 se detalla de forma específica los datos fuente utilizados para la creación de esta dimensión, además de una breve descripción de cada atributo que la conforma.

Tabla 43. Mapeo Dimensión Inscrito

Nombre de la estructura del <i>Datamart</i>		Área de inscripciones y admisiones
Dimensión		dim_inscrito
Atributos	Fuente	Descripción
id_inscrito	hadm_aspirante.numero_documento	Identificador Aspirante
tipo_documento	hadm_aspirante.tipo_documento	Documento de Identificación
nombres	hadm_aspirante.nombres	Nombres del Aspirante
apellidos	hadm_aspirante.apellidos	Apellidos del Aspirante
fecha_nacimiento	hadm_aspirante.fecha_nacimiento	Fecha de Nacimiento del Aspirante
Sexo	hadm_aspirante.sexo	Sexo del Aspirante
Edad	calculada	Edad del Aspirante
estado_civil	hadm_aspirante.estado_civil	Estado Civil del Aspirante
tipo_sangre	a_inscripcion.tipo_sangre	Tipo de Sangre del Aspirante
Estrato	hadm_aspirante.estrato1	Estrato del Aspirante
Cupo	especial.detalle_especial1	Cupo del Estudiante
validacion_icfes	hadm_aspirante.validacion_icfes	Aspirante Valido ICFES
victima_conflicto	a_inscripcion.victima_conflicto	Aspirante es Víctima de Conflicto
Etnia	etnia.etnia_descr	Etnia que pertenece Aspirante
capacidad	capacidades.capac_descr	Capacidad del Aspirante
discapacidad	discapacidad.discap_descr	Discapacidad del Aspirante
tipo_ingreso	a_inscripcion.tipo_ingreso	Tipo de Ingreso del Aspirante

Fuente. Esta investigación.

Los datos de la dimensión inscrito se deben mapear de la siguiente manera:

- **id_inscrito.** El atributo almacenará el número de identificación del aspirante.

- **tipo_documento.**El atributo almacenará el tipo de documento del aspirante, los datos aceptados para el atributo son: CEDULA DE CIUDADANIA, TARJETA DE IDENTIDAD, PASAPORTE, CEDULA EXTRANJERA, NUMERO UNICO DE IDENTIFICACION, para los aspirantes que no se conozca el tipo de documento se deberá ingresar el valor DESCONOCIDO.
- **nombres.** El atributo almacenará los dos nombres del aspirante, el valor debe estar en mayúscula.
- **apellidos.** El atributo almacenará los dos apellidos del aspirante, el valor debe estar en mayúscula.
- **fecha_nacimiento.** El atributo almacenará la fecha de nacimiento del aspirante, el formato para el ingreso de la fecha es mes/día/año, por ejemplo: 03/27/1992.
- **sexo.** El atributo almacenará el género del aspirante, los datos aceptados son: M, para sexo masculino, F, para sexo femenino.
- **edad.** El atributo almacenará la edad del aspirante, como se observa en la tabla 43 el atributo no tiene una fuente de datos definida de las tablas transaccionales. El valor del atributo se ingresa con un procedimiento en cual involucra la fecha de nacimiento y el año de ingreso del aspirante.
- **estado_civil.** El atributo almacenará el estado civil del aspirante, los datos aceptados para el atributo son: CASADO, MADRE SOLTERA, RELIGIOSO, SEPARADO, SOLTERO UNION LIBRE, VIUDO. Para los aspirantes que no se conozca su estado civil se deberá ingresar el valor DESCONOCIDO.
- **tipo_sangre.** El atributo almacenará el tipo de sangre al que pertenece el aspirante, los datos aceptados para el atributo son: A+, A-, AB+, AB-, B+, B-, O+, O-. Para los aspirantes que no se conozca el tipo de sangre se deberá ingresar el valor DESCONOCIDO.
- **estrato.** El atributo almacenará el estrato el cual pertenece el aspirante, los datos aceptado para el atributo son de 0 a 10.Para los aspirantes que no se conozca su estrato se deberá ingresar el valor DESCONOCIDO.
- **cupo.** El atributo almacenará el tipo de cupo por el cual el aspirante ingresa, los datos aceptados para el atributo son: CUPO REGULAR, DEPORTISTA DESTACADO, DESPLAZADOS NARIÑO-PUTUMAYO, DISCAPACITADOS, EXTRANJERO SIN NACIONALIDAD COLOMBIANA, HIJO DE VICTIMA DE SECUESTRO, HIJO DE EMPLEADO, HIJO DE VICTIMA DE DESAPARICION, MEJOR BACHILLER DEPARTAMENT DE NARIÑO, MUNICIPIOS DEPRIMIDOS SOCIO-ECONOMICAMENTE, NEGRITUDES DE LA ZONA PACIFICA NARIÑENSE, PERTENECIENTES

A CABILDOS INDIGENAS DE NARIÑO, PERTENECIENTE CABILDO INDIGENA PUTUMAYO, PROFESIONALES, REINSERTADOS.

- **validacion_icfes.** EL ATRIBUTO ALMACENARÁ información que permita conocer si el aspirante valido el ICFES, los datos aceptados para el atributo son SI, NO. Para los aspirantes que no se conozca esta información se deberá ingresar el valor DESCONOCIDO.
- **victima_conflicto.** El atributo almacenará información que permita conocer si el aspirante es víctima de conflicto, los datos aceptados para el atributo son SI, NO. Para los aspirantes que no se conozca esta información se deberá ingresar el valor DESCONOCIDO.
- **etnia.** El atributo almacenará la etnia que pertenece el aspirante, los datos aceptados para el atributo son: ACHAGUA, AWA(CUAIKER), COFÁN O KOFÁN, CHIRICOA, EMBERA CAITO O KATÍO, EMBERA CHAMI, EMPERARA SIAPIDARA, GUAMBIANO, IGNA, KAMSA O KAMENTSÁ, NASA (PAÉZ), NEGRITUDES, PASTOS, PIAPOCO (DZASE), PIAROA, PIRATAPUTO, QUILLACINGAS, WITOTO, YANACONA, YUKUNA, ZENÚ. Para los aspirantes que no pertenezcan a ninguna etnia se deberá ingresar el valor NO APLICA, además a los aspirantes que no se conozcan su etnia está el valor DESCONOCIDO.
- **capacidad.** El atributo almacenará la capacidad del aspirante, los datos aceptados para el atributo son: CON TALENTO CIENTIFICO, CON TALENTO SUBJETIVO, CON TALENTO TECNOLOGICO, SUPERDOTADO. Para los aspirantes que no posean ninguna capacidad se deberá ingresar el valor NO APLICA, además para los aspirantes que no se conozca esta información se encuentra el valor DESCONOCIDO.
- **discapacidad.** El atributo almacenará la discapacidad del aspirante, los datos aceptados para el atributo son: AUTISMO, BAJA VISION DIAGNOSTICA, CEGUERA, DEFICIENCIA COGNITIVA (RETARDO MENTAL), HIPOACUSIA A BAJA AUDICION, LESION NEUROMUSCULAR, MULTIPLE, PARALISIS CEREBRAL, SINDROME DE DOWN, SORDERA PROFUNDA. Para los aspirantes que no posean ninguna discapacidad se deberá ingresar el valor NO APLICA, además para los aspirantes que no se conozca esta información se encuentra el valor DESCONOCIDO.
- **tipo_ingreso.** El atributo almacenará el tipo de ingreso del aspirante, los datos aceptados para el atributo son: CAMBIO DE SECCION, REINGRESO, TRANSFERENCIA, TRASLADO, NO INFORMA. Para los aspirantes que no se conozca el tipo de ingreso se deberá ingresar el valor DESCONOCIDO.

- **Dimensión admitido.** En la tabla 44 se detalla de una forma específica los datos fuente utilizados para la creación de la dimensión admitido además de una breve descripción de cada atributo que la conforma.

Los datos de la dimensión admitido se deben mapear de la siguiente forma:

- **id_admitido.** El atributo almacenará el código que permitirá identificar si el aspirante es admitido.
- **admitido.** El atributo almacenará el detalle del estado del estudiante, los datos aceptados para el atributo son: ADMITIDO, NO ADMITIDO.

Tabla 44. Mapeo dimensión admitido

Nombre de la estructura del <i>Datamart</i>		Área de inscripciones y admisiones
Dimensión		dim_admitido
Atributos	Fuente	Descripción
id_admitido	Esta Investigación	Identificador Admitido
Admitido	Esta Investigación	Descripción Admitido

Fuente. Esta investigación.

3.4.1.3 Mapeo matrículas y población estudiantil. Para esta área se describe la tabla de hechos “fact_matriculados” y las dimensiones relacionadas, de acuerdo a los siguientes puntos:

- **Tabla de hechos matrículas y población estudiantil.** En la tabla 45 se describen las posibles operaciones que se pueden realizar sobre la métrica o medida de esta tabla de hechos. La medida identificada debe ser numérica, los datos correctos deben mapearse de la siguiente forma:

- **num_alumnos.** La métrica almacenará la cantidad de estudiantes, el único valor aceptado es 1.

- **Dimensión vigente.** En la tabla 46 se detalla de una forma específica los datos fuente utilizados para la creación de la dimensión vigente además de una breve descripción de cada atributo que la conforma.

Tabla 45. Operaciones de la tabla de hechos matrículas y población estudiantil

Tabla de hechos	fact_matriculados				
Atributo 1	num_alumnos				
Dimensiones	Suma	Conteo	Promedio	Minimo	Máximo
dim_alumno	Si	Si			
dim_carrera	Si	Si			
dim_estado	Si	Si			
dim_region	Si	Si			
dim_tiempo	Si	Si			

Fuente. Esta investigación.

Tabla 46. Mapeo dimensión vigente

Nombre de la estructura del <i>Datamart</i>	Área de matrículas y población estudiantil	
Dimensión	dim_vigente	
Atributos	Fuente	Descripción
id_vigente	Esta investigación	Identificador semestre
detalle_vigente	Esta investigación	Detalle vigente

Fuente. Esta investigación.

Los datos para la dimensión vigente deben mapearse de la siguiente forma:

- **id_vigente.** El atributo almacenara el código que permitirá identificar si el estudiante se encuentra matriculado o no, o si canceló el semestre o año (para la carrera de derecho).

- **detalle_vigente.** El atributo almacenará el detalle del estado del estudiante, los datos aceptados para el atributo son: MATRICULADO, NO MATRICULADO, MATRICULA CANCELADA.
- **Dimensión semestre.** En la tabla 47 se detalla de una forma específica los datos fuente utilizados para la creación de la dimensión semestre además de una breve descripción de cada atributo que la conforma.

Tabla 47. Mapeo dimensión semestre

Nombre de la estructura del <i>Datamart</i>		Área de matrículas y población estudiantil
Dimensión		dim_semestre
Atributos	Fuente	Descripción
id_semestre	Esta investigación	Identificador semestre
semestre	Esta investigación	Detalle semestre

Fuente. Esta investigación.

- **id_semestre.** el atributo almacenará el código que permitirá identificar el semestre al pertenecen los estudiantes en un determinado periodo o año (para la carrera de derecho).
- **semestre.** el atributo almacenará el detalle del semestre, los datos aceptados para el atributo son: 1 SEMESTRE, 2 SEMESTRE, 3 SEMESTRE, 4 SEMESTRE, 5 SEMESTRE, 6 SEMESTRE, 7 SEMESTRE, 8 SEMESTRE, 9 SEMESTRE, 10 SEMESTRE, INFORMACIÓN MATERIAS ADICIONALES/INTERNADO ROTATORIO, INFORMACIÓN EGRESADOS, INFORMACIÓN GRADUADOS, INFORMACIÓN DESCONOCIDA.

3.4.1.4 Mapeo de materias y rendimiento académico .Para esta área se describe las tablas de hechos “fact_rendacad”, “fact_eventos”, “fact_eventos_mat” y “fact_cancelaciones”, junto con las dimensiones relacionadas, de acuerdo a los siguientes puntos:

- **Tabla de hechos rendimiento académico y materias matriculadas.** En la tabla 48 se describe las posibles operaciones que se pueden realizar sobre cada

una de las métricas o medidas de la presente tabla de hechos rendimiento académico y materias matriculadas.

Tabla 48. Operaciones de la tabla de hechos rendimiento académico y materias matriculadas

Tabla de hechos	fact_rendacad				
Atributo 1	nota				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_alumno			Si	Si	Si
dim_materia			Si	Si	Si
dim_carrera			Si	Si	Si
dim_estado			Si	Si	Si
dim_region			Si	Si	Si
dim_colegio			Si	Si	Si
dim_aprobacion			Si	Si	Si
dim_tiempo			Si	Si	Si
Atributo 2	num_matriculadas				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_alumno	Si	Si			
dim_materia	Si	Si			
dim_carrera	Si	Si			
dim_estado	Si	Si			
dim_region	Si	Si			
dim_colegio	Si	Si			
dim_aprobacion	Si	Si			
dim_tiempo	Si	Si			

Fuente. Esta investigación.

Las medidas en la presente tabla de hechos deben ser valores numéricos, los datos correctos deben mapearse de la siguiente forma:

- **nota.** La métrica almacenará la nota definitiva por materia matriculada de los estudiantes, los datos aceptados para el atributo son valores numéricos de coma flotante.
- **num_matriculadas.** La métrica almacenará la cantidad de materias matriculadas, el único valor aceptado es 1.

Tabla 49. Operaciones de la tabla de hechos eventos sobre materias

Tabla de hechos	fact_eventos_mat				
Atributo 1	Nota				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_alumno			Si	Si	Si
dim_materia			Si	Si	Si
dim_carrera			Si	Si	Si
dim_eventos_mat			Si		
dim_estado			Si	Si	Si
dim_region			Si	Si	Si
dim_colegio			Si	Si	Si
dim_aprobacion			Si	Si	Si
dim_tiempo			Si	Si	Si
Atributo 2	num_eventosmat				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_alumno	Si	Si			
dim_materia	Si	Si			
dim_carrera	Si	Si			
dim_eventos_mat	Si	Si			
dim_estado	Si	Si			
dim_region	Si	Si			
dim_colegio	Si	Si			
dim_aprobacion	Si	Si			
dim_tiempo	Si	Si			

Fuente. Esta investigación.

- **Tabla de hechos eventos sobre materias.** En la tabla 49 se describe las posibles operaciones que se pueden realizar sobre cada una de las métricas o medidas de la tabla de hechos eventos sobre materias.

Las medidas en la tabla de hechos eventos sobre materias deben ser valores numéricos, los datos correctos deben mapearse de la siguiente forma:

- **nota.** La métrica almacenará la nota definitiva por eventos de materias registrados en los estudiantes, los datos aceptados para el atributo son valores numéricos de coma flotante.
- **num_eventosmat.** La métrica almacenará el número de veces que se registra un evento en las materias, el único valor aceptado es 1.

- **Tabla de hechos de eventos sobre estudiantes.** En la tabla 50 se describe las posibles operaciones que se pueden realizar sobre cada una de las métricas o medidas de la tabla de hechos cancelaciones.

Tabla 50. Operaciones de la tabla de hechos de eventos sobre estudiantes.

Tabla de hechos	fact_eventos				
Atributo 1	num_eventos				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_alumno	Si	Si			
dim_eventos	Si	Si			
dim_carrera	Si	Si			
dim_estado	Si	Si			
dim_colegio	Si	Si			
dim_region	Si	Si			
dim_tiempo	Si	Si			

Fuente. Esta investigación.

- **num_eventos.** La métrica almacenará el número de veces que se registra un evento sobre los estudiantes, el único valor aceptado es 1.

• **Tabla de hechos cancelaciones.** En la tabla 51 se describe las posibles operaciones que se pueden realizar sobre cada una de las métricas o medidas de esta tabla de hechos. Las medida identificada debe ser numérica, los datos correctos deben mapearse de la siguiente forma:

- **num_canceladas.** La métrica almacenará el número de veces que se ha cancelado las materias, el único valor aceptado es 1.

Tabla 51. Operaciones de la tabla de hechos cancelaciones

Tabla de hechos	fact_cancelaciones				
Atributo 1	num_canceladas				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_alumno	Si	Si			
dim_materia	Si	Si			
dim_carrera	Si	Si			
dim_region	Si	Si			
dim_tiempo	Si	Si			

Fuente. Esta investigación.

Tabla 52. Mapeo Dimensión aprobación

Nombre de la estructura del <i>Datamart</i>	Área de Matrículas de materias y rendimiento académico	
Dimensión	dim_aprobacion	
Atributos	Fuente	Descripción
id_aprobacion	Esta investigación	Identificador aprobación
detalle_aprobacion	Esta investigación	Si estudiante aprobó materias

Fuente. Esta investigación.

- **Dimensión Aprobación.** En la tabla 52 se detalla de una forma específica los datos fuente utilizados para la creación de esta dimensión, además de una breve descripción de cada atributo que la conforma.

Los datos de la dimensión aprobación se deben mapear de la siguiente forma:

- **id_aprobacion:** El atributo almacenará el número de identificación de aprobación.
- **detalle_aprobación.** El atributo almacenará el estado de aprobación de las materias, los datos aceptados para el atributo son: APROBADAS, REPROBADAS.

- **Dimensión estado.** En la tabla 53 se detalle de forma específica los datos fuente utilizados para la creación de la dimensión estado, además de una breve descripción de cada atributo que lo conforma.

Tabla 53. Mapeo dimensión estado

Nombre de la estructura del <i>Datamart</i>		Área de Matrículas de materias y rendimiento académico
Dimensión		dim_estado
Atributos	Fuente	Descripción
id_estado	Esta investigación	Identificador estado del estudiante
detalle_estado	Esta investigación	Descripción estado estudiante

Fuente. Esta investigación.

Los datos de la dimensión admitido se deben mapear de la siguiente forma:

- **id_estado.** El atributo almacenará el número de identificación del estado del alumno.
 - **detalle_estado.** El atributo almacenará el estado académico del alumno. Los datos aceptados para el atributo son: NO GRADUADOS, EGRESADOS, GRADUADOS.
- **Dimensión eventos.** En la tabla 54 se detalle de forma específica los datos fuente utilizados para la creación de esta dimensión, además de una breve

descripción de cada atributo que lo conforma. Los datos se deben mapear de la siguiente forma:

- **id_evento.** El atributo almacenará el número de identificación del evento.
- **nombre_evento.** el atributo almacenará la información de eventos estudiantiles, tales como: MATRICULA ANULADA, SEMESTRE CANCELADO, REINGRESO, MATRICULA DE HONOR, RETIRO POR FUERZA MAYOR, TRASLADO, entre otros, el valor de los datos se almacenará en letras mayúsculas.

Tabla 54. Mapeo dimensión eventos

Nombre de la estructura del <i>Datamart</i>		Área de Matrículas de materias y rendimiento académico
Dimensión		dim_eventos
Atributos	Fuente	Descripción
id_evento	acumula_def.cod_materia	Identificador Evento
nombre_evento	materias.detalle_largo	Descripción evento

Fuente. Esta investigación.

- **Dimensión eventos materias.** En la tabla 55 se detalle de forma específica los datos fuente utilizados para la creación de esta dimensión.

Tabla 55. Mapeo dimensión eventos materias

Nombre de la estructura del <i>Datamart</i>		Área de Matrículas de materias y rendimiento académico
Dimensión		dim_eventos_mat
Atributos	Fuente	Descripción
id_convencion	Esta Investigación	Identificador Evento Materias
detalle_convencion	Esta Investigación	Descripción del Evento

Fuente. Esta investigación.

- **id_convencion.** El atributo almacenará el número de identificación de la convención utilizada para los eventos sobre las materias.
- **detalle_convencion.** El atributo almacenará las convenciones de eventos que se registran sobre las materias. Los datos aceptados para el atributo son: CE: CURSOS ESPECIALES, CV: CURSOS DE VACACIONES, CO: MATERIAS HOMOLAGADAS, V: VALIDACIONES.

3.4.1.5 Mapeo profesores. Para esta área se describe las tablas de hechos “fact_profesores” y “fact_evaluaciones”, junto con las dimensiones relacionadas, de acuerdo a los siguientes puntos:

- **Tabla de hechos profesores.** En la tabla 56 se describe las posibles operaciones que se pueden realizar sobre cada una de las métricas o medidas de la tabla de hechos profesores.

Tabla 56. Operaciones de la tabla de hechos profesores

Tabla de hechos	fact_profesores				
Atributo 1	num_profesores				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_profesor	SI	SI			
dim_carrera	SI	SI			
dim_tiempo	SI	SI			
dim_nivel_academico	SI	SI			
dim_vinculacion	SI	SI			

Fuente. Esta investigación.

La medida en la tabla de hechos profesores deben ser un valor numérico, los datos correctos deben mapearse de la siguiente forma:

- **num_profesores.** La métrica permitirá conocer el número de profesores, el único valor aceptado es 1.
- **Tabla de hechos evaluaciones.** En la tabla 57 se describe las posibles operaciones que se pueden realizar sobre cada una de las métricas o medidas de

la tabla de hechos evaluaciones. Las medidas en la tabla de hechos evaluación deben ser valores numéricos, los datos correctos deben mapearse de la siguiente forma:

- **gran_fortaleza.** La métrica permitirá conocer número de profesores que obtuvieron la calificación gran fortaleza en la pregunta determinada. Los únicos datos aceptados son (0 y 1).
- **fortaleza.** La métrica permitirá conocer número de profesores que obtuvieron la calificación fortaleza en la pregunta determinada. Los únicos datos aceptados son (0 y 1).
- **transicion.** La métrica permitirá conocer número de profesores que obtuvieron la calificación transición en la pregunta determinada. Los únicos datos aceptados son (0 y 1).
- **debilidad.** La métrica permitirá conocer número de profesores que obtuvieron la calificación debilidad en la pregunta determinada. Los únicos datos aceptados son (0 y 1).
- **gran_debilidad.** La métrica permitirá conocer número de profesores que obtuvieron la calificación gran debilidad en la pregunta determinada. Los únicos datos aceptados son (0 y 1).

Tabla 57. Operaciones de la tabla de hechos evaluaciones

Tabla de hechos	fact_evaluacion				
Atributo 1	gran_fortaleza				
Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_profesor	Si				
dim_carrera	Si				
dim_tiempo	Si				
dim_pregunta	Si				
dim_nivel_academico	Si				
dim_vinculacion	Si				
Atributo 2	fortaleza				

Tabla 57. Continuación

Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_profesor	Si				
dim_carrera	Si				
dim_tiempo	Si				
dim_pregunta	Si				
dim_nivel_academico	Si				
dim_vinculacion	Si				

Atributo 3	transicion
-------------------	------------

Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_profesor	Si				
dim_carrera	Si				
dim_tiempo	Si				
dim_pregunta	Si				
dim_nivel_academico	Si				
dim_vinculacion	Si				

Atributo 4	debilidad
-------------------	-----------

Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_profesor	Si				
dim_carrera	Si				
dim_tiempo	Si				
dim_pregunta	Si				
dim_nivel_academico	Si				
dim_vinculacion	Si				

Atributo 5	gran_debilidad
-------------------	----------------

Dimensiones	Suma	Conteo	Promedio	Mínimo	Máximo
dim_profesor	Si				
dim_carrera	Si				
dim_tiempo	Si				
dim_pregunta	Si				
dim_nivel_academico	Si				
dim_vinculacion	Si				

Fuente. Esta investigación.

- **Dimensión profesor.** En la tabla 58 se detalla de una forma específica los datos fuente utilizados para creación de la dimensión profesor además de una breve descripción de cada atributo que la conforma.

Tabla 58. Mapeo dimensión profesor

Nombre de la estructura del <i>Datamart</i>		Área de Profesores
Dimensión		dim_profesor
Atributos	Fuente	Descripción
id_profesor	profesores.cedula	Identificador Profesor
nombres	profesores.nombres	Nombres Profesores
apellidos	profesores.apellidos	Apellidos Profesores
sexo	profesores.sexo	Sexo del Profesor
titulo	profesores.titulo	Título del Profesor

Fuente. Esta investigación.

Los datos de la dimensión profesores se deben mapear de la siguiente forma:

- **id_profesor.** El atributo almacenará el número de identificación del profesor.
- **nombres.** El atributo almacenará los dos nombres del profesor, el valor debe estar en mayúscula.
- **apellidos.** El atributo almacenará los dos apellidos del profesor, el valor debe estar en mayúscula.
- **sexo.** El atributo almacenará el género del profesor, los datos aceptados son (MASCULINO, FEMENINO).
- **titulo.** El atributo almacenará el título obtenido por el profesor, el valor debe estar en mayúscula.

- **Dimensión nivel académico.** En la tabla 59 se detalla de una forma específica los datos fuente utilizados para creación de la dimensión nivel académico además de una breve descripción de cada atributo que la conforma. Los datos de la dimensión nivel académico se deben mapear de la siguiente forma:

- **id_nivel_academico.** El atributo almacenará el código de identificación del nivel académico del profesor.
- **nivel_academico.** El atributo almacenará el nivel de formación académica del profesor, los datos aceptados para el atributo son (PROFESIONAL, ESPECIALISTA, MAGISTER, DOCTOR).

Tabla 59. Mapeo dimensión nivel académico

Nombre de la estructura del <i>Datamart</i>		Área de Profesores
Dimensión		dim_nivel_academico
Atributos	Fuente	Descripción
id_nivel_academico	Esta Investigación	Identificador Nivel Académico
nivel_academico	profesores.cod_nivel_academico	Nivel Académico

Fuente. Esta investigación.

- **Dimensión vinculación.** En la tabla 60 se detalla de una forma específica los datos fuente utilizados para creación de la dimensión vinculación además de una breve descripción de cada atributo que la conforma.

Tabla 60. Mapeo dimensión vinculación

Nombre de la estructura del <i>Datamart</i>		Área de Profesores
Dimensión		dim_vinculacion
Atributos	Fuente	Descripción
id_vinculacion	vinculaciones.cod_vin	Identificador vinculación
Vinculación	vinculaciones.vinculacion	Tipo de Vinculación

Fuente. Esta investigación.

- **id_vinculacion.** El atributo almacenará el código de identificación de la vinculación del profesor.

- **vinculacion.** El atributo almacenará el tipo de vinculación del profesor con la universidad, los datos aceptados para el atributo son (HORA CATEDRA, HORA POR SERVICIOS PRESTADOS, HORA CATEDRA HONORARIOS, HORA CATEDRA JUBILADO, INVITADO, TIEMPO COMPLETO OCASIONAL, TIEMPO COMPLETO).
- **Dimensión pregunta.** En la tabla 61 se detalla de una forma específica los datos fuente utilizados para creación de la dimensión pregunta además de una breve descripción de cada atributo que la conforma.

Tabla 61. Mapeo dimensión pregunta

Nombre de la estructura del <i>Datamart</i>		Área de Profesores
Dimensión		dim_pregunta
Atributos	Fuente	Descripción
id_pregunta	Esta Investigación	Identificador de la pregunta
formulario	formularios.nom_form	Detalle Formulario
caracteristica	caracteristicas.nom_car	Detalle Característica
variable	variables.nom_var	Detalle Pregunta

Fuente. Esta investigación.

- **id_pregunta.** El atributo almacenará el código de la pregunta en el formulario de evaluación.
- **formulario.** El atributo almacenará la descripción del formulario, el valor debe estar en mayúscula.
- **caracteristica.** El atributo almacenará la descripción de la caracteristica, el valor debe estar en mayúscula.
- **variable.** El atributo almacenará la descripción de la pregunta, el valor debe estar en mayúscula.

3.4.2 Back Room.

Es el área del Data Warehouse responsable de extraer y preparar los datos. En esta sección se explicara de manera general el proceso de extracción, transformación y carga de los datos al prototipo funcional.

Para el proceso se parte de los datos fuentes de los sistemas de información de la Universidad de Nariño. Una de las políticas del Data Warehouse es no modificar los sistemas de la organización pues se estarían alterando sus procesos de negocios y de esta forma los procesos OLTP, por tal razón el centro de informática brindo una copia de seguridad de las bases de datos que conformaban el área académica de la Universidad de Nariño.

- **Extracción.** La Universidad de Nariño maneja varias bases de datos en sus sistemas de información. En el proyecto se hizo una migración de las tablas importantes para el desarrollo del prototipo.

Para la extracción correcta de las tablas necesarias en la construcción del prototipo se crearon una serie de scripts que hicieran la labor de seleccionar los registros y guardarlos en tablas temporales, por ejemplo para la dimensión inscrito se toman valores de las tablas a_inscripcion, hadm_aspirante, hadm_academicos, capacidades, discapacidad, etnia. Pero no son necesarios todos los registros de estas tablas. Por tal razón se creó la tabla temporal tmp_inscrito.

- **Transformación.** Para la transformación de los datos se hizo el mapeo descrito en la sección 3.4.1, las tablas fuentes tenían muchos problemas de consistencia en sus registros por tal razón fue necesario hacer la limpieza, para entender mejor la transformación se puede observar la estructura de los scripts que se encuentran en los anexos del presente documento. Además para dicha limpieza se hizo uso de la herramienta EXDACLET, desarrollada en los laboratorios del grupo de investigación GRIAS de programa de ingeniería de sistemas de la Universidad de Nariño. Las tablas temporales no tienen una estructura exactamente igual a las fuentes pues solo se hizo extracción de los datos que interesaban, de esta forma se hicieron las transformaciones.

- **Carga.** Luego de tener los datos transformados en las tablas temporales y los scripts listos, se hace el proceso de carga en cada una de las dimensiones y tablas de hechos que conforman el diseño del Datamart. Este paso consiste en tomar estos datos fuentes y guardarlos en el cubo de tal forma que queden listos para que se puedan utilizar herramientas OLAP o de análisis multidimensional.

3.4.3 Front Room. El *Datamart* de información académica de la Universidad de Nariño está estructurado de forma que se puede ver la información de manera multidimensional de las áreas de inscripciones y admisiones, área de matrículas y población estudiantil, área de materias y rendimiento académico, área de profesores con respecto a los alumnos, programas y profesores clasificados por semestre y años. El formato de visualización de los datos se los detallara en el capítulo 4.

4. PRUEBAS Y RESULTADOS

Para las pruebas y resultados se realizó un prototipo funcional, denominado *datamartudenar*, construido a partir de los esquemas dimensionales diseñados en el capítulo 3, bajo el SGBD *PostgreSQL*. El análisis multidimensional de los datos almacenados en este prototipo se realizó bajo la herramienta *STARCUBE*, creada en los laboratorios KDD del grupo de investigación GRIAS del programa de ingeniería de sistemas de la Universidad de Nariño. A continuación se describen en detalle la estructura del prototipo *datamartudenar* la instalación y el manejo de la herramienta *STARCUBE* sobre el área de inscripciones y admisiones, además de realizar un breve análisis de resultados sobre la información obtenida a partir del análisis multidimensional de las diferentes áreas del *Datamart*.

4.1 PROTOTIPO DEL *DATAMART* DE INFORMACIÓN ACADÉMICA.

El prototipo surge del modelo en constelación diseñado en la fase de análisis multidimensional, y sobre este se puede decir lo siguiente:

- La implementación está desarrollada bajo el SGBD *PostgreSQL*, en su versión 8.4.4-1, bajo el sistema operativo Microsoft Windows 7. Para el análisis multidimensional se adoptó la herramienta *STARCUBE*.
- Dentro de la base de datos utilizada para implementar el modelo dimensional diseñado existen tres esquemas: El esquema académico: el cual contiene las fuentes de datos pertenecientes al área de inscripciones y admisiones, el área de matrículas y población estudiantil y el área de materias y rendimiento académico. El esquema profesores: el cual contiene las tablas que soportan la información del área de profesores. El esquema público el cual contiene las dimensiones y las tablas de hechos que se crearon a partir del modelo multidimensional planteado, y que además se alimentan de los datos almacenados dentro de las tablas en esquemas mencionados anteriormente, los scripts relacionados con este proceso se describen en el anexo B de este documento.
- La información que se tiene en cuenta son datos proporcionados por el Centro de Informática de la Universidad de Nariño, a partir del año 2003, en cada periodo (A o B) o año (para el caso de derecho), de todas las facultades para las carreras: LICENCIATURA EN INFORMÁTICA, LICENCIATURA EN MATEMÁTICAS,

LICENCIATURA EN FILOSOFÍA Y LETRAS, PSICOLOGÍA, INGENIERÍA AGRONÓMICA, INGENIERÍA DE SISTEMAS, INGENIERÍA AGROINDUSTRIAL, ECONOMÍA, ADMINISTRACIÓN DE EMPRESAS, DERECHO, ZOOTECNIA, TECNOLOGÍA EN PROMOCIÓN DE LA SALUD, MEDICINA, MEDICINA VETERINARIA, INGENIERÍA AGROFORESTAL, DISEÑO GRÁFICO Y ARQUITECTURA. Se tienen en cuenta las sedes y las jornadas a las que pertenecen cada una de las carreras mencionadas anteriormente.

4.2 HERRAMIENTA DE ANÁLISIS MULTIDIMENSIONAL STARCUBE

STARCUBE es un software de apoyo a toma de decisiones, mediante la creación de cubos de datos y su manipulación.

STARCUBE permite realizar todas las operaciones de OLAP como son *dice*, *slice*, *dril down*, *roll up*, *pivot*, todas estas operaciones para cambiar la perspectiva del análisis de los cubos de datos.

STARCUBE es una herramienta desarrollada bajo el lenguaje de programación java.

Para el correcto funcionamiento de la herramienta se necesita tener instalado el siguiente software:

- *PostgreSQL* 8.3 o superior.
- La Máquina Virtual de Java jre6 o superior.
- Usuario del SGBD admin que tenga permiso de creación de base de datos, super usuario, puede crear roles.
- Crear y restaurar la base de datos Starcube con la utilización de archivo StarCube.sql.

Para el correcto funcionamiento de la herramienta *STARCUBE* y el prototipo creado los requerimientos mínimos de hardware son:

- Procesador Core 2 duo 1.86 GHz.
- Memoria RAM 2gb.
- Disco Duro de 200 Gb.

4.2.1 Instalación STARCUBE. Para este proceso es necesario descomprimir el archivo starcube.rar y ejecutar el archivo StarCube.jar dentro de la carpeta base de la herramienta.

- **Inicio de Sesión en STARCUBE.** Para iniciar sesión en *STARCUBE* es necesario tener un rol con todos los permisos de superusuario, en este caso se tiene en cuenta el rol admin. La figura 40 muestra la pantalla de inicio de sesión de STARCUBE.

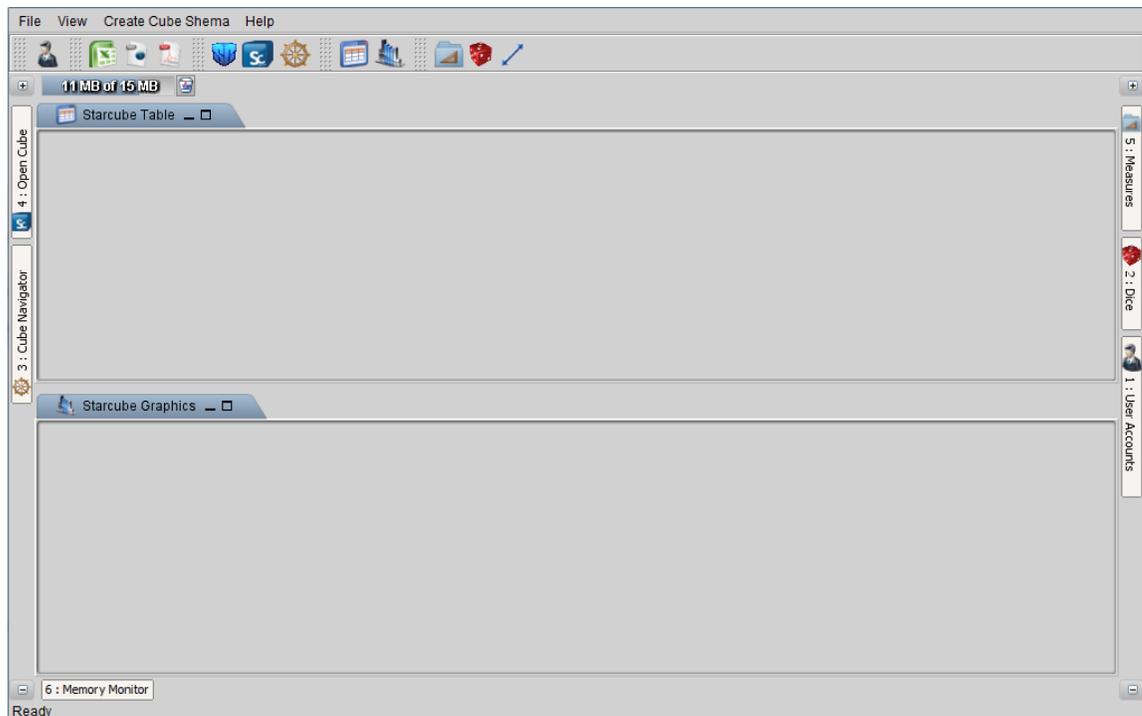
Figura 40. Inicio de sesión STARCUBE



Fuente. Esta investigación.

En figura 41 muestra la pantalla inicial, en la parte superior se observa el menú principal que contiene una serie de botones que permiten aplicar las operaciones: *roll up*, *drilldown*, *slice*, *dice*. Además también este menú permite exportar los resultados de las consultas en diferentes formatos.

Figura 41. Pantalla inicial STARCUBE



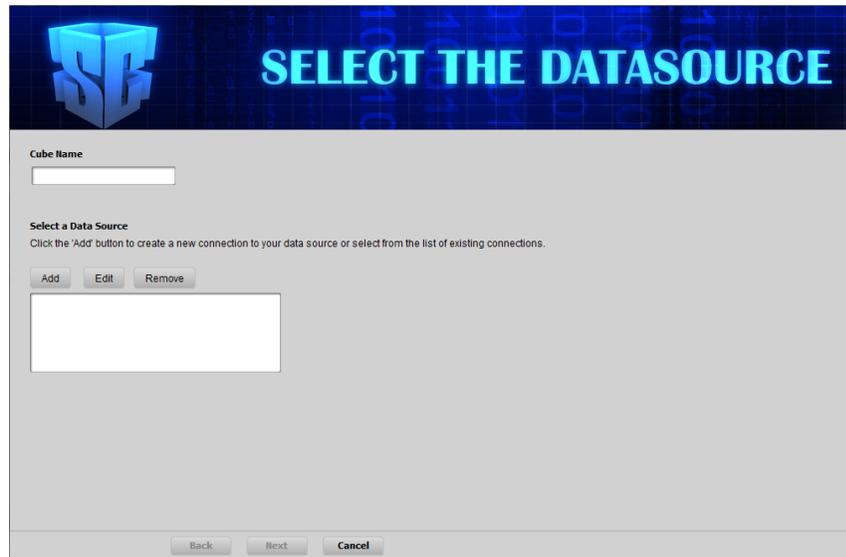
Fuente. Esta investigación.

4.3 CUBO DE DATOS DEL ÁREA DE INSCRIPCIONES Y ADMISIONES

Para entender la creación de un cubo de datos, en esta sección se describe la creación del cubo de datos del área de inscripciones y admisiones. El procedimiento se realiza de la siguiente forma:

- **Crear el esquema de datos.** Para realizar este paso se utilizó el botón  el cual abrirá otra ventana. En esta ventana será posible empezar a construir un Cubo de datos. Para comenzar, se debe escoger la base de datos de origen y escribir el nombre que se desee para el Cubo en el campo de texto denominado “*Cube Name*”. El esquema del cubo será guardado en un archivo con formato XML.

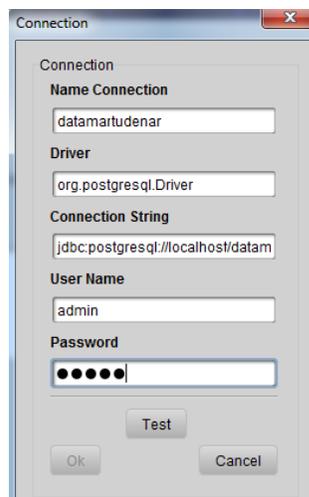
Figura 42. Creación de un cubo de datos



Fuente. Esta investigación.

En la figura 42 se observa en el cuadro de texto superior el nombre del cubo el cual se llamará cobertura. En la parte inferior se observa la fuente de datos o la conexión a la base de datos, en este caso se añadirá una nueva presionando el botón "Add". La figura 43 muestra la información de la conexión que se realiza para el cubo de datos de inscripciones y admisiones.

Figura 43. Añadir / Editar una conexión.

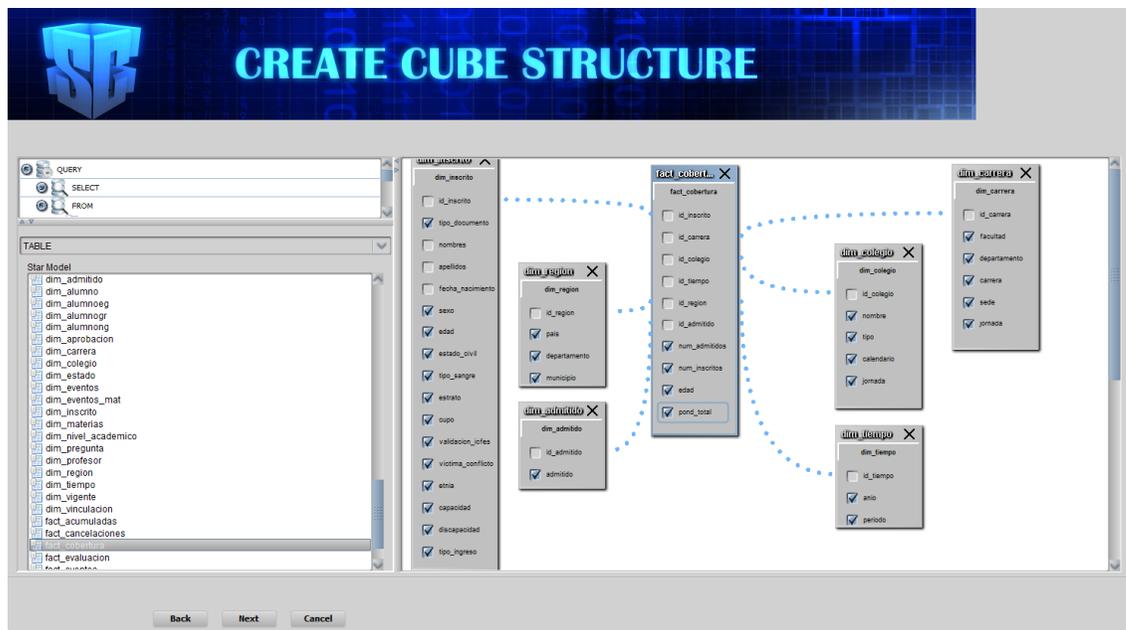


Fuente. Esta investigación.

Para verificar si la conexión se realizó con éxito se debe presionar el botón “Test”. Aparecerá una alerta diciendo si la conexión fue exitosa o no. Para finalizar la conexión se presiona el botón OK. Si la conexión a la base de datos se realiza más de una vez, solo se crea una vez, en este caso la conexión datamartudenar servirá para crear los cubos pertenecientes a las diferentes áreas identificadas.

- **Creación de la estructura del cubo.** Para crear el esquema del cubo se necesita arrastrar la tabla de hechos y las tablas de dimensiones al panel central de trabajo, tratando de que se visualice tal como se describe en el modelo dimensional del área correspondiente. En la figura 44 se muestra la estructura del cubo de datos de inscripciones y admisiones.

Figura 44. Creación de la estructura cubo de datos inscripciones y admisiones.



Fuente. Esta investigación.

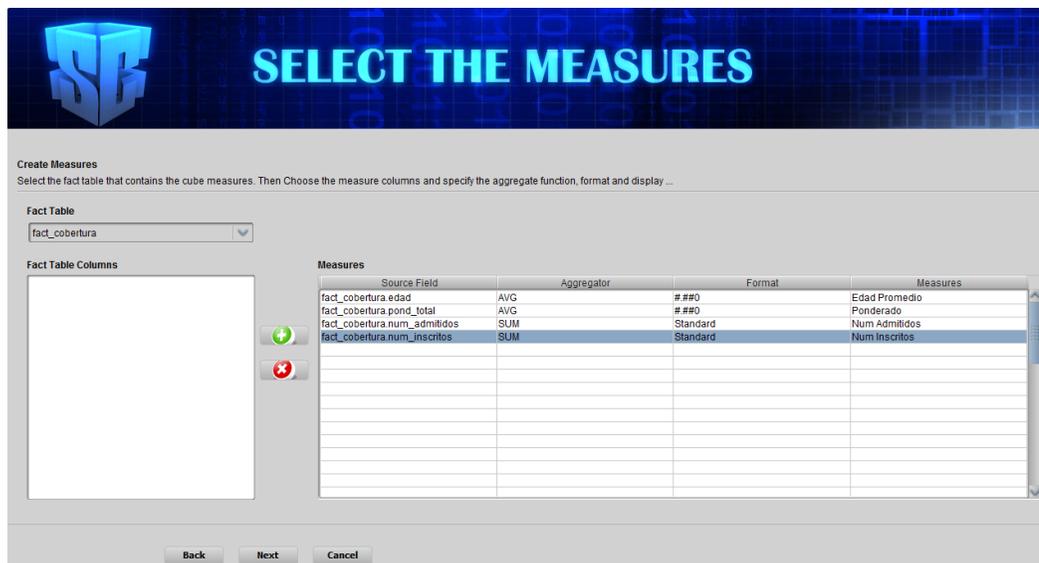
Los enlaces entre las tabla de hechos y las dimensiones se realizan seleccionando la clave relacionada con la dimensión y mediante un *Click*sostenido se arrastra hacia la clave primaria homónima de la dimensión. Por ejemplo: se selecciona el atributo id_carrera, de la tabla de hechos “fact_cobertura” y mediante un *Click*sostenido se lo relaciona con la clave primaria de la dimensión carrera “dim_carrera”, el cual es su homónimo id_carrera.

Para seleccionar los campos a analizar se debe hacer *Click* sobre las casillas de verificación de cada tabla de hechos o de dimensión. Para las tablas de hechos se seleccionan únicamente las medidas y para las dimensiones los valores de descripción.

Para el siguiente paso se debe presionar el botón “*Next*” y a continuación se elegirá las medidas o métricas.

- **Selección de las medidas o métricas.** Las métricas son los valores sobre los cuales se hará las operaciones de agregación en las consultas. Para seleccionar las medidas, se debe dar *Click* en la lista desplegable y seleccionar la tabla de hechos o *FactTable*, luego cuando aparezcan los campos debajo, se escogen los valores sobre los cuales se realizará el análisis multidimensional, haciendo *Click* en el botón . Las métricas aparecerán en la rejilla del lado derecho. Sobre las medidas seleccionadas se harán operaciones tales como sumas, conteos, promedios, máximos o mínimos. Los resultados de este paso pueden verse en la figura 45. En el caso del área de inscripciones y admisiones se seleccionan las métricas num_admitidos, num_inscritos, edad y ponderado, sobre las cuales se realiza operaciones de suma para el número de admitidos y el número de inscritos y promedio para la edad y el ponderado de matrícula.

Figura 45. Selección de medidas



Fuente. Esta investigación.

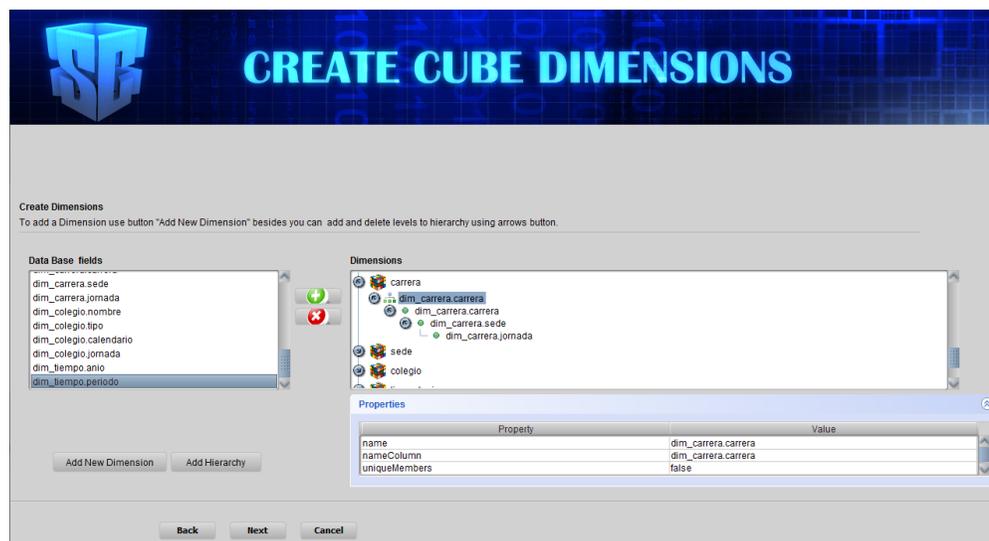
El siguiente paso es la creación de dimensiones y jerarquías.

- **Creación de las dimensiones y jerarquías.** Para crear una dimensión se selecciona un atributo de dimensión, dando *Click* en el botón “*Add New Dimension*”. De esta forma se crea la dimensión y automáticamente una jerarquía y un nivel con el mismo nombre, o el nombre que se quiera. Se tiene la opción de añadir nuevas jerarquías, como por ejemplo el caso de ciudad, la cual iniciaría con país, luego con departamento y por último el municipio. Esto se realiza con el botón “*AddHierarchy*”, o un nuevo nivel a las jerarquías con el botón .

Para remover cualquier dimensión, jerarquía o nivel, se utiliza el botón .

Al crear las dimensiones con las jerarquías y niveles se tendrá algo como se muestra en la figura 46.

Figura 46. Agregar dimensiones y jerarquías.



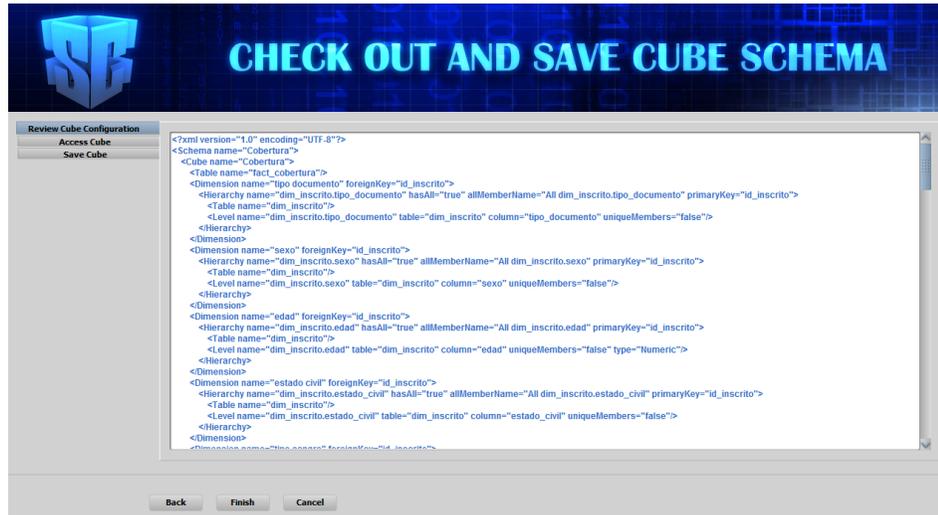
Fuente. Esta investigación.

En este paso para las dimensiones de inscripciones y admitidos se selecciona las dimensiones con sus jerarquías y niveles correspondientes. Como por ejemplo en la dim_carrera se selecciona la facultad, añadiendo las jerarquías de departamento y carrera, además de los niveles de jornada y sede, siendo todo este conjunto una sola dimensión del cubo.

El siguiente paso es la revisión de esquema del cubo.

- **Revisión del esquema del cubo (XML).** Al finalizar la creación del cubo, se podrá revisar la configuración de este esquema mediante un archivo XML que se visualiza en la pantalla final de la creación de cubos en la pestaña “Review Cube Configuration” como se muestra en la figura 47. Si no se ha generado el código XML es porque hubo errores en los pasos anteriores.

Figura 47. Esquema XML del cubo de datos.



Fuente. Esta investigación.

Finalmente cuando el esquema del cubo se haya configurado adecuadamente, se guarda haciendo *Click* en el botón “Save Cube” de la pestaña *Save Cube*, en la pantalla final. El esquema del Cubo se almacena en la base de datos *PostgreSQL*, que administra los datos que se manejan en *STARCUBE*. El resultado final se muestra en la figura 48.

Figura 48. Almacenamiento de un Cubo.



Fuente. Esta investigación.

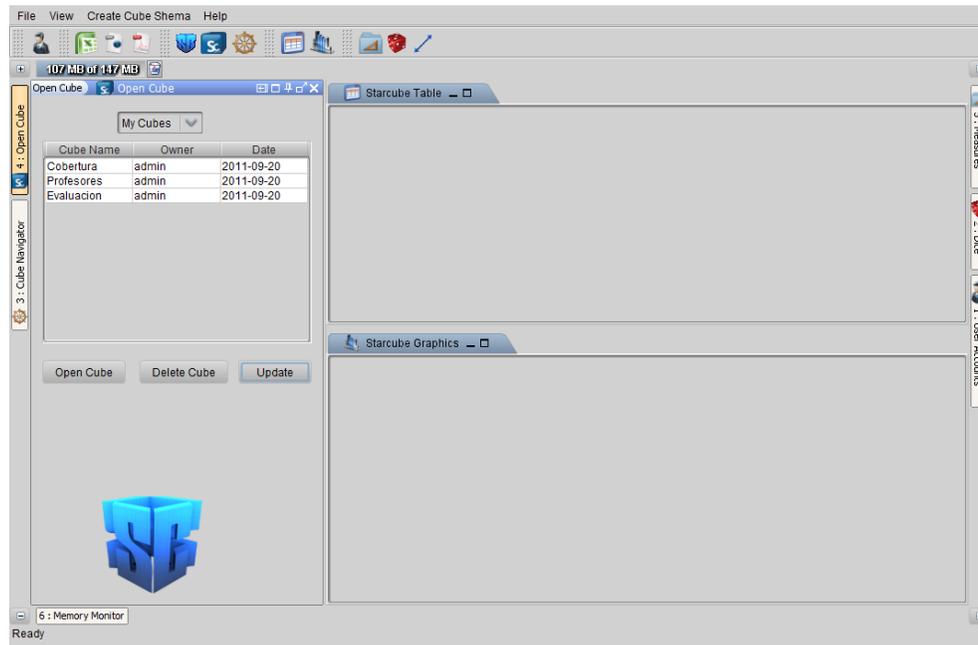
4.4 MANEJO DE LA HERRAMIENTA STARCUBE SOBRE EL ÁREA DE INSCRIPCIONES Y ADMISIONES

La explicación del menú principal se realizara de izquierda a derecha:

-  Crear, Eliminar y Editar usuarios de STARCUBE.
-  Exportar los resultados de la consulta a un archivo xls.
-  Exportar los resultados de la consulta en un archivo html.
-  Exportar los resultados de la consulta a un archivo pdf.
-  Crear un nuevo cubo, este proceso esta detallado en la sección 4.3.
-  Permite abrir o eliminar un cubo ya creado, En este panel están listados todos los cubos creados. En la lista desplegable hay dos opciones que puede

seleccionar: “My Cubes” donde se encuentran todos los cubos creados y “Shared Cubes” donde están los cubos públicos, compartidos por otros usuarios.

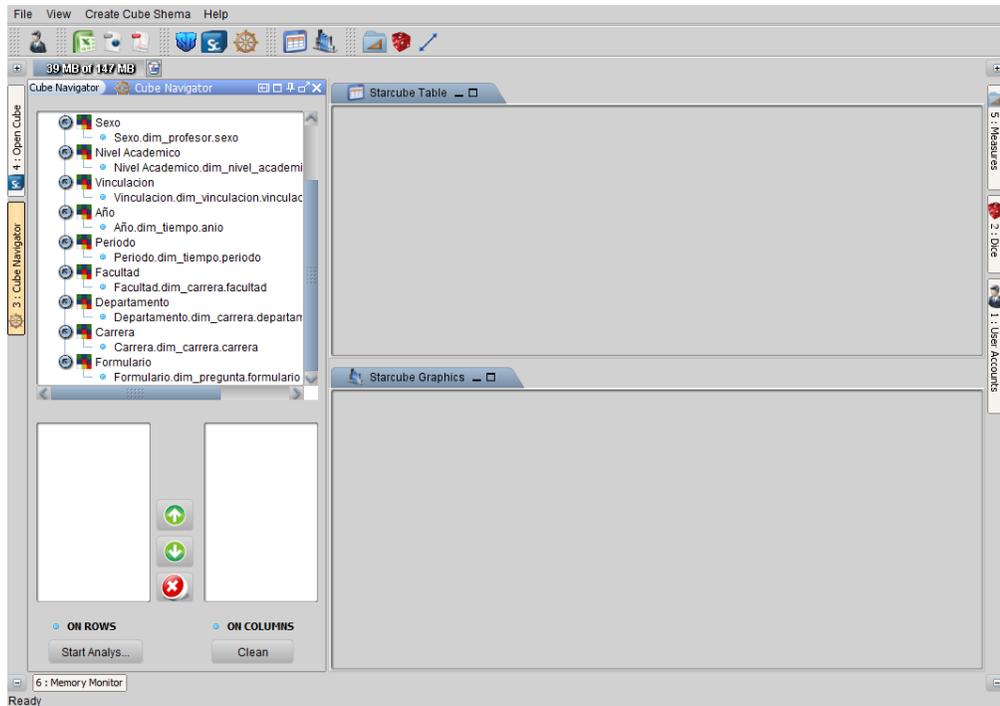
Figura 49. Cubos creados STARCUBE



Fuente. Esta investigación.

-  Permite navegar sobre un cubo abierto. En este panel, la información de un Cubo de datos es representada en una estructura de árbol, mostrando el nombre del Cubo, las Medidas (incluidas en una sola Jerarquía) y las diferentes Dimensiones con sus Jerarquías.
-  Permite visualizar la consulta en una tabla OLAP para un análisis intuitivo de los datos arrojados, en la figura 51 se muestra resultados de analizar las medidas añadidas anteriormente clasificadas por año en el área de inscripciones y admisiones.
-  Permite visualizar gráficamente los resultados de la consulta, en la figura 52 se muestra los resultados de la consulta en una representación tipo pastel de cada una de las medidas.

Figura 50. Navegación sobre un cubo OLAP.



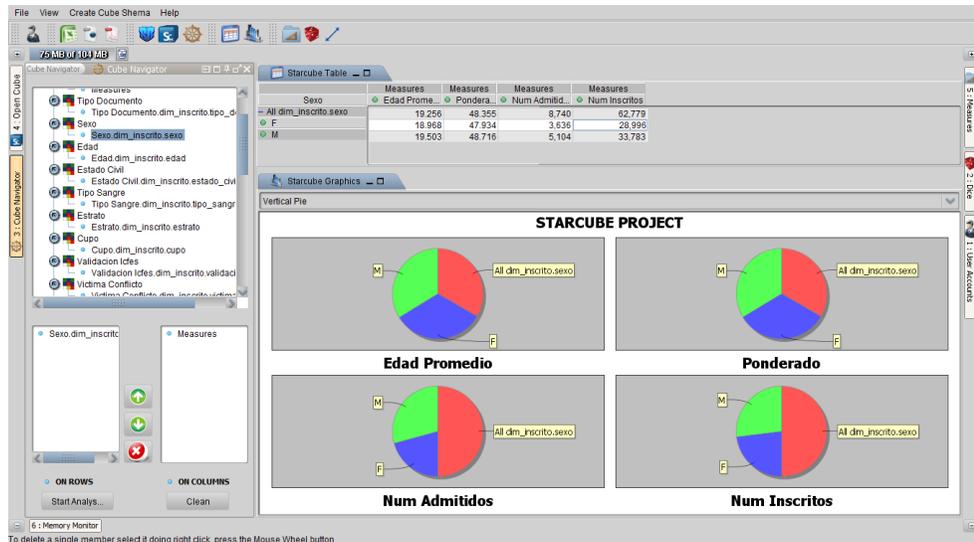
Fuente. Esta investigación.

Figura 51. Tabla OLAP

	Measures	Measures	Measures	Measures
Año	Edad Promedio	Ponderado	Num Admitidos	Num Inscritos
- All dim_tiempo.anio	19.256	48.355	8,740	62,779
+ 2003	19.039	47.275	925	6,190
+ 2004	19.164	48.555	1,364	9,392
+ 2005	19.126	49.168	1,054	7,921
+ 2006	19.209	49.875	1,116	8,018
+ 2007	19.301	46.221	1,073	8,545
+ 2008	19.237	47.729	961	7,746
+ 2009	18.96	48.239	1,165	6,908
+ 2010	20.227	48.149	558	3,443
+ 2011	19.703	50.693	524	4,616

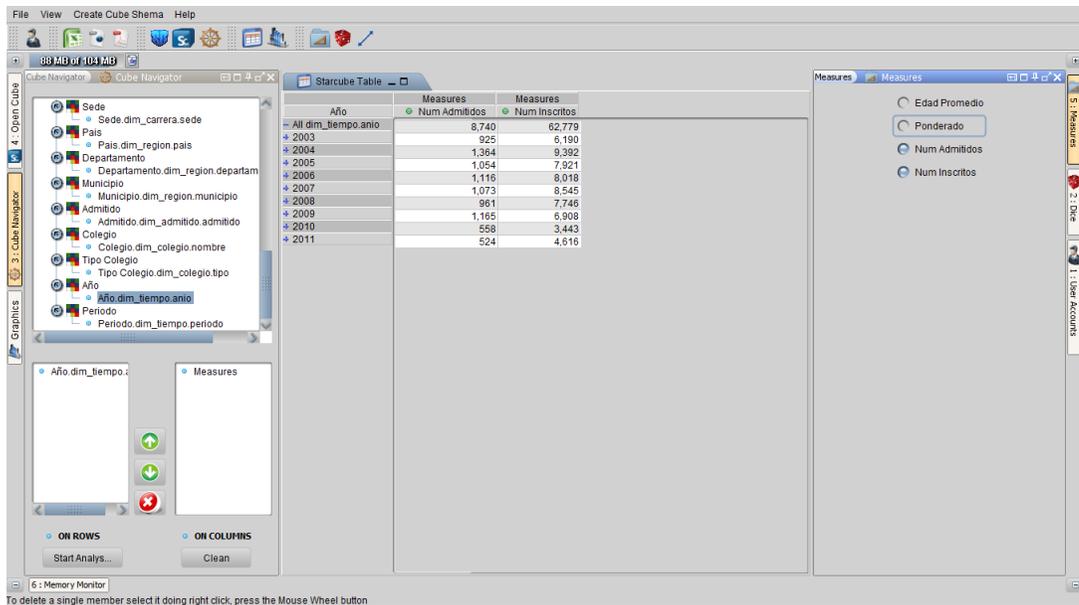
Fuente. Esta investigación.

Figura 52. Diagrama tipo pastel inscripciones y admisiones



Fuente. Esta investigación.

Figura 53. Restricción medidas o métricas.



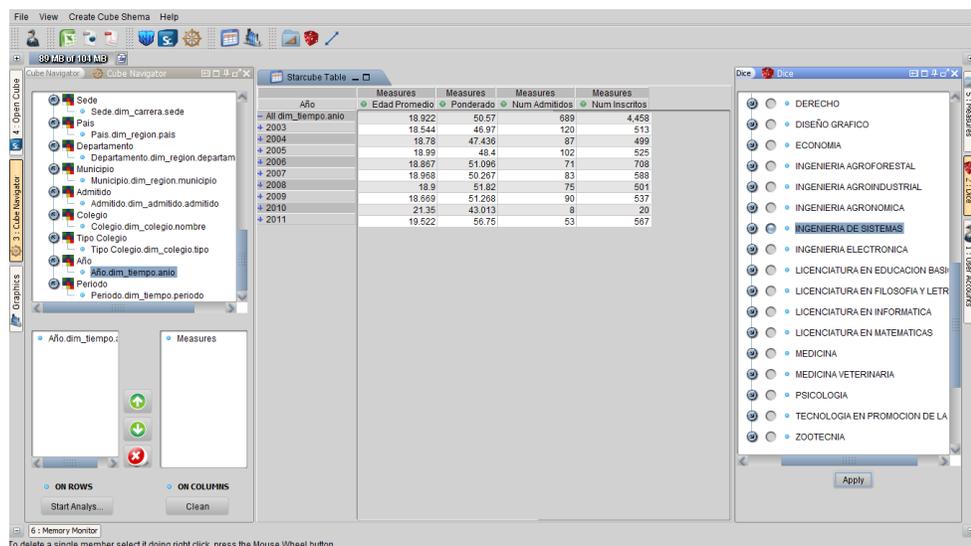
Fuente. Esta investigación.

- 
 Permite añadir o eliminar una Medida de la Tabla OLAP, seleccione cuáles Medidas quiere o no, mostrar en la Tabla, haciendo clic en las casillas de verificación. En la figura 53 se restringe las medidas edad_promedio y ponderado,

se puede ver que la tabla central muestra solo las medidas de num_admitidos y num_inscritos, clasificados por año.

- 
 Permite hacer la operación Dice, en el panel seleccione cuál Miembro será la restricción para su consulta OLAP.

Figura 54. Operación *dice*

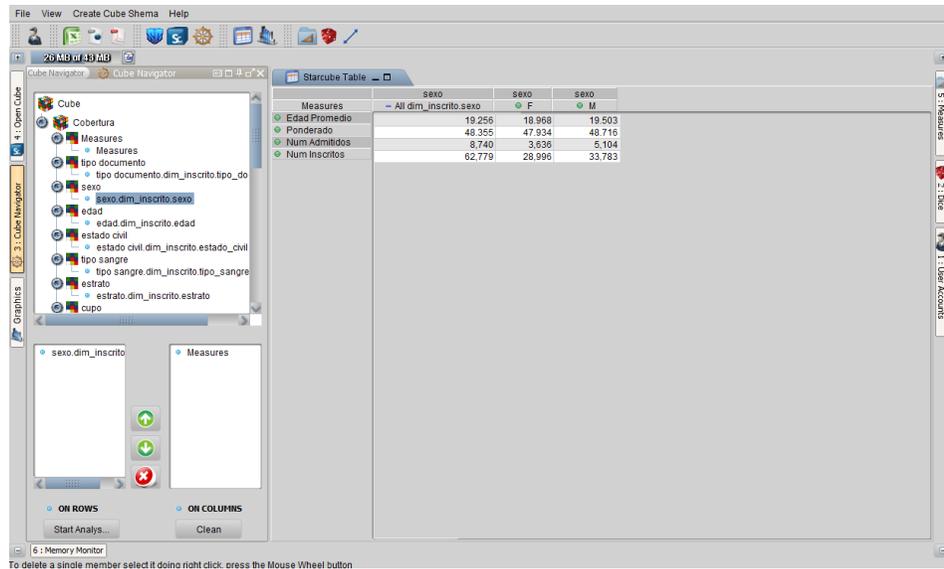


Fuente. Esta investigación.

En la figura 54 se visualizan los datos únicamente de INGENIERÍA DE SISTEMAS, teniendo en cuenta las medidas num_inscritos, num_admitidos, edad y ponderado, clasificando estos valores por año. Para hacer esta operación se debe tener al menos una dimensión fuera del análisis actual, es decir, que no se encuentre en la tabla OLAP.

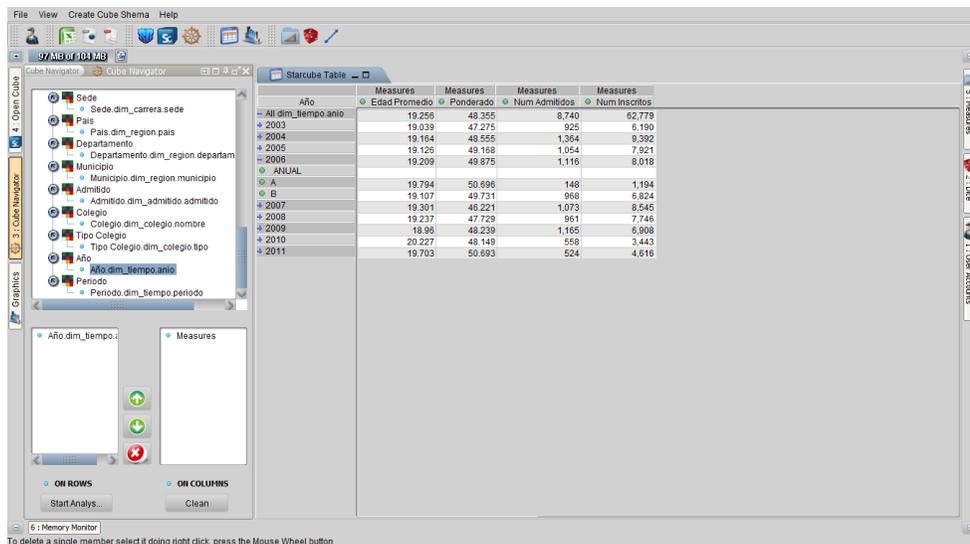
- 
 Permite aplicar la operación pivote a la tabla OLAP en la figura 55 se puede observar el cambio de eje de la tabla.
- 
 Operación *Drill Down*, esto modificará la Tabla OLAP mostrando los hijos que contiene el miembro seleccionado. La figura 56 describe esta operación.

Figura 55. Operación *pivote*



Fuente. Esta investigación.

Figura 56. Operación *drilldown*



Fuente. Esta investigación.

4.5 ANÁLISIS DE RESULTADOS

Para el análisis de resultados se hicieron pruebas en la herramienta STARCUBE, creando los cubos de datos de cada área como se hizo en la sección anterior, haciendo el correspondiente análisis multidimensional de cada área sobre el prototipo mencionado en la sección 4.1. Los resultados se describen en los siguientes párrafos.

4.5.1 Información inscripciones y admisiones. A continuación se hace un análisis sobre una muestra de información académica, la cual permite realizar análisis de las medidas el número de inscritos, número de admitidos, la edad promedio de los aspirantes y el ponderado de matrículas sobre el cubo de inscripciones y admisiones. Esta información se obtiene del prototipo que se construyó en este proyecto, el cual soporta el diseño planteado.

Tabla 62. Información de inscripciones detallada por año

Año	Edad Promedio	Ponderado Matricula	Número de Admitidos	Número de Inscritos
Total	19.256	48.355	8,740	62,779
2003	19.039	47.275	925	6,19
2004	19.164	48.555	1,364	9,392
2005	19.126	49.168	1,054	7,921
2006	19.209	49.875	1,116	8,018
A	19.794	50.696	148	1,194
B	19.107	49.731	968	6,824
2007	19.301	46.221	1,073	8,545
2008	19.237	47.729	961	7,746
2009	18.961	48.239	1,165	6,908
2010	20.227	48.149	558	3,443
2011	19.703	50.693	524	4,616

Fuente. Esta investigación.

En la tabla 62 se observa que desde el periodo A del 2003 hasta el periodo A de 2011 se inscribieron 62779 personas en la Universidad de Nariño, de los cuales fueron admitidos 8740 personas, que corresponden a menos del 20% de la población de aspirantes. Además en el año 2006 en el periodo A se puede detallar (Mediante la operación *drilldown*) que se inscribieron 1194 frente a los 6824 del periodo B, esto responde a que en la Universidad de Nariño la mayoría de los

programas se abren en el periodo B de cada año (Esto hasta el periodo B de 2010).

En las siguiente tabla se analizan las mismas medidas mencionadas en el análisis de la tabla 62, esta vez para el programa de Psicología, lo cual responde a la operación Dice dentro del cubo de datos, la cual restringe los datos únicamente para el programa mencionado. La tabla 63 responde a los resultados obtenidos al realizar esta operación.

Tabla 63. Información de inscripciones y admisiones del programa de psicología

Año	Edad Promedio	Ponderado Matricula	Número de Admitidos	Número de Inscritos
Total	19.338	34.145	497	4,897
2003	18.850	47.207	41	492
2004	19.136	48.092	149	698
2005	19.263	39.878	48	639
2006	19.380	27.582	49	545
2007	19.422	25.889	48	645
2008	19.355	26.923	55	563
2009	19.097	28.445	53	580
2010	24.857	50.424	0	14
2011	19.897	28.905	54	721

Fuente. Esta investigación.

En la tabla 63 se puede ver que para el programa de Psicología se han inscrito a partir del año 2003 hasta 2011, 4897 personas de las cuales se admitieron 497, lo cual representa una tasa de absorción de tan solo el 10%. El año con más aspirantes fue el 2011 con 721 personas, mientras que el año con más aspirantes admitidos fue 2004 con 149 personas.

En la tabla 64 se describe la información que permite analizar las métricas mencionadas anteriormente, pero esta vez no por año sino en los programas utilizados en el prototipo del *datamart*. Se observa que carreras son las que representan más interés por parte de los aspirantes. La carrera con más aspirantes entre el año 2003 y 2011, fue el programa de DERECHO con 8477 personas, valor que representa el 13% del total de inscritos, mientras que el programa que tiene menos interés es LICENCIATURA EN MATEMÁTICAS, con

1482 personas inscritas, representando un 2.3% del total de aspirantes en los diferentes programas.

Tabla 64. Información de inscripciones y admisiones por programas

Carrera	Edad Promedio	Ponderado Matricula	Número de Admitidos	Número de Inscritos
Total	19.256	48.355	8,740	62,779
ADMINISTRACION DE EMPRESAS	19.174	49.871	824	5,668
ARQUITECTURA	19.216	39.615	380	2,610
DERECHO	19.616	53.423	747	8,477
DISEÑO GRAFICO	19.333	30.163	365	2,529
ECONOMIA	19.138	51.593	476	3,302
INGENIERIA AGROFORESTAL	19.676	49.612	427	2,756
INGENIERIA AGROINDUSTRIAL	19.470	50.707	396	3,305
INGENIERIA AGRONOMICA	19.238	48.507	488	2,349
INGENIERIA DE SISTEMAS	18.922	50.570	689	4,458
INGENIERIA ELECTRONICA	18.725	53.003	693	4,045
LICENCIATURA EN FILOSOFIA Y LETRAS	19.534	49.281	414	1,524
LICENCIATURA EN INFORMATICA	19.320	49.191	378	1,924
LICENCIATURA EN MATEMATICAS	18.935	51.136	386	1,482
MEDICINA	18.988	53.935	372	4,257
MEDICINA VETERINARIA	19.568	49.857	419	3,178
PSICOLOGIA	19.338	34.145	497	4,897
TECNOLOGIA EN PROMOCION DE LA SALUD	19.079	48.337	407	3,587
ZOOTECNIA	19.178	48.478	382	2,431

Fuente. Esta investigación.

En la siguiente tabla se expresa la información de aspirantes cuya ciudad de procedencia es PASTO e IPIALES.

Tabla 65. Información de inscripciones y admisiones de Pasto e Ipiiales

Municipio	Edad Promedio	Ponderado Matricula	Número de Admitidos	Número de Inscritos
Total	19.256	48.355	8,740	62,779
IPIALES	19.070	47.021	698	4,501
PASTO	19.100	48.949	5,193	35,935

Fuente. Esta investigación.

En la tabla 65 se observa que de los 62799 inscritos entre los años 2003 y 2011, 35935 son de la ciudad de Pasto.

4.5.2 Información matrículas y población estudiantil.A continuación se hace un análisis sobre una muestra de información académica, la cual permite hacer análisis de población estudiantil, teniendo en cuenta la medida de número de alumnos sobre el cubo de matrículas y población estudiantil. Esta información se obtiene del prototipo que se construyó en este proyecto, el cual soporta el diseño planteado en el capítulo 3.

Tabla 66. Población estudiantil por año

Año	Sexo	Número de alumnos
		Total
2003	Total 2003	1,346
2004	Total 2004	4,061
2005	Total 2005	6,749
2006	Total 2006	9,315
	F	4,101
	M	5,214
-ANUAL-	Total anual 2006	439
	F	214
	M	225
A	Total periodo A	3,886
	F	1,739
	M	2,147
B	Total periodo B	4,990

Tabla 66. Continuación.

	F	2,148
	M	2,842
2007	Total 2007	11,835
2008	Total 2008	14,023
2009	Total 2009	16,164
2010	Total 2010	18,412

Fuente. Esta investigación.

En la tabla 66 se observa que el total de alumnos en el año 2010 de la muestra tomada para el prototipo, mencionada en la sección 4.1, es de 18412. La suma de la población de todos los años corresponde a 81905 registros de alumnos. Para el año 2006 el total de alumnos es de 9315 alumnos, 4101 mujeres y 5214 hombres, en el periodo A de este año se registraron 3886 alumnos y en las carreras anuales (DERECHO) se registraron 439 personas, de las cuales 214 son mujeres y 225 hombres.

En la siguiente tabla se analiza la misma medida de población, esta vez presentando en detalle el estado de los estudiantes del programa de Ingeniería de Sistemas en el año 2010, haciendo la operación dice, la cual restringe los valores de la muestra tomados.

Tabla 67. Información de estado de estudiantes de Ingeniería de Sistemas.

Año	Estado del estudiante	Número de alumnos
		Total población
2003	Total 2003	149
2004	Total 2004	389
2005	Total 2005	592
2006	Total 2006	790
2007	Total 2007	970
2008	Total 2008	1,148
2009	Total 2009	1,362
2010	Total 2010	1,564
	EGRESADOS 2010	65
	GRADUADOS 2010	42
	NO GRADUADOS 2010	1,457

Tabla 67. Continuación.

-ANUAL-	ANUAL 2010	0
A	TOTAL PERIODO A 2010	782
	EGRESADOS A 2010	31
	GRADUADOS A 2010	10
	NO GRADUADOS A 2010	741
B	TOTAL PERIODO B 2010	782
	EGRESADOS B 2010	34
	GRADUADOS B 2010	32
	NO GRADUADOS B 201	716

Fuente. Esta investigación.

En la tabla 67 se observa que para el año 2010 hay un total de 1564 alumnos de los cuales 65 son egresados, 42 son graduados y 1457 aún no se han graduado. Al detallar los periodos se observa que el periodo anual tiene el valor de 0 registros, esto es porque Ingeniería de sistemas es una carrera semestral. Para el periodo A de 2010 hay 782 alumnos, de los cuales 31 estudiantes son egresados, 10 graduados y 741 no graduados. Respecto al periodo B los valores de la población total no varían, esto es porque Ingeniería de sistemas no se abrió para este periodo. Se puede observar que en el 2010 hay 31 egresados y en 2010 34, además en el periodo B existen 32 estudiantes graduados, durante el periodo A y B de 2010 se graduaron 22.

En la siguiente tabla se analiza el número de estudiantes en ingeniería de sistemas en primer semestre, detallando el estado de matrícula, para el 2008.

Tabla 68. Información matriculas ingeniería de sistemas primer semestre

Año	Vigencia estudiantes	Número de alumnos
		Total
2003	Total 2003	149
2004	Total 2004	389
2005	Total 2005	592
2006	Total 2006	790
2007	Total 2007	970
2008	Total 2008	1,148
	MATRICULA CANCELADA	15
	MATRICULADO	552

Tabla 68. Continuación.

	NO MATRICULADO	581
-ANUAL-	Total Anual	0
A	Total A 2008	530
	MATRICULA CANCELADA A 2008	10
	MATRICULADO A 2008	246
	NO MATRICULADO A 2008	274
B	Total B 2008	618
	MATRICULA CANCELADA B 2008	5
	MATRICULADO B 2008	306
	NO MATRICULADO B 2008	307
2009	Total 2009	1,362
2010	Total 2010	1,564

Fuente. Esta investigación.

En la tabla 68 se observa que para año 2008 existían 1148 alumnos de los cuales 15 cancelaron matrícula, 552 se matricularon y 581 no se matricularon, estos valores corresponden al 1.31%, 48.08% y 50.61% respectivamente del total de la muestra tomada para el programa de Ingeniería de Sistemas.

4.5.3 Información de materias y rendimiento académico.A continuación se hace un análisis de la muestra tomada para el prototipo, la cual permite hacer análisis sobre las materias y el rendimiento académico.

Tabla 69. Notas y materias matriculadas por año

Año	Promedio de Notas	Materias Matriculadas
Total	3.654	341,253
2003	3.567	9,257
2004	3.643	24,099
2005	3.638	36,617
2006	3.622	44,373
2007	3.647	49,959
2008	3.667	51,335
-ANUAL-	3.991	6,309
A	3.643	21,714
B	3.601	23,312
2009	3.65	52,968

Tabla 69. Continuación.

-ANUAL-	3.989	6,452
A	3.687	21,580
B	3.531	24,936
2010	3.69	48,967
2011	3.704	23,678

Fuente. Esta investigación.

En la tabla 69 se puede observar que el promedio de notas en el año 2008 es de 3.667, en 51335 materias matriculadas en los diferentes programas que hacen parte de la muestra tomada para el prototipo. Para el periodo A de 2008 el promedio es de 3.643 y en el periodo A de 2010 es de 3.687, comparando los dos periodos A de 2008 y 2009 se puede decir que el rendimiento académico en general es casi igual, ya que no varía en mucho respecto a las materias matriculadas y el promedio de materias general, de las carreras que se ofrecen semestralmente.

En la siguiente tabla se muestra el número de eventos de estudiantes a partir del año 2003.

Tabla 70. Información de eventos sobre estudiantes.

Eventos	Número de eventos
Total de Eventos	7179.0
ANULA INSCRIPCION PARA REINGRESO	3.0
ANULADA MATRICULA	5.0
ANULADA MATRICULA POR NO CUMPLIR CON LOS REQUISITOS EXIGIDOS	3.0
CAMBIO DE ESPECIALIDAD	
CAMBIO DE SECCION	51.0
CANCELACION DE SEMESTRE Y RESERVA DE CUPO POR FUERZA MAYOR	1.0
CANCELADO SEMESTRE Y RESERVA DE CUPO	136.0
MATRICULA ANULADA	5.0
MATRICULA COMO ESTUDIANTE ESPECIAL	2.0
MATRICULA DE HONOR	2169.0
MATRICULADO SIN ASIGNATURAS A CURSAR	1.0

Tabla 70. Continuación

NO REGISTRO ASIGNATURAS - PAGO LIQUIDACION	6.0
PERDIDA DEFINITIVA DE DERECHO A CONTINUAR ESTUDIOS	233.0
PERDIDA TEMPORAL DE DERECHO A CONTINUAR ESTUDIOS	1061.0
REINGRESO	876.0
REINGRESO POR SANCION DE I SEMESTRE	11.0
REINGRESO Y TRASLADO	1.0
RESERVA DE CUPO	5.0
RESERVA DE CUPO POR FUERZA MAYOR	9.0
RESERVA DE CUPO POR RETIRO VOLUNTARIO	6.0
RETIRO POR FUERZA MAYOR	417.0
RETIRO VOLUNTARIO	15.0
RETIRO VOLUNTARIO Y DEFINITIVO DEL PROGRAMA	113.0
SANCION POR 30 DIAS	2.0
SANCION POR ACUERDO 051 DE 2003 CONSEJO SUPERIOR ARTICULO 84	1414.0
SANCION: EXPULSION DE LA UNIVERSIDAD DE NARIÑO	1.0
SEMESTRE CANCELADO	176.0
TRANSFERENCIA	22.0
TRASLADO	428.0
TRASLADO ANULADO	1.0

Fuente. Esta investigación.

En la tabla 70 se observa que el evento los eventos que más se han dado son las matrículas de honor y las pérdidas temporales de estudio desde el año 2003 hasta el 2011 en toda la muestra tomada para el prototipo.

En la siguiente tabla se muestra la información de eventos especiales sobre materias clasificados por año para los estudiantes no graduados clasificados por año:

Tabla 71. Información de eventos sobre materias

Año	Convención	Promedio Eventos	Número de eventos
	Total	3.725	10,908
2003	Total 2003	3.016	185
2004	Total 2004	3.858	587
2005	Total 2005	3.638	680
	CE: CURSO ESPECIAL 2005	3.263	19
	CO: HOMOLOGADA 2005	4.091	422
	CV: CURSO DE VACACIONES 2005	2.856	227
	V: VALIDACION 2005	3.05	12
2006	Total 2006	3.588	892
2007	Total 2007	3.748	1,340
2008	Total 2008	3.868	1,236
2009	Total 2009	3.731	1,690
2010	Total 2010	3.728	2,265
	CE: CURSO ESPECIAL 2010	3.439	955
	CO: HOMOLOGADA 2010	3.975	1,171
	CV: CURSO DE VACACIONES 2010	0	0
	V: VALIDACION 2010	3.638	139
2011	Total 2011	3.73	2,033

Fuente. Esta investigación.

En la tabla 71 se observa que desde el año 2003 hasta el año 2011 se han dado 10908 eventos especiales sobre las materias, con un promedio total de 3.725. Para el año 2005 el promedio fue de 3.588 en un total de 892 eventos especiales, se observa que en para este año se realizaron 19 cursos especiales, con un promedio de 3.263. Para el año 2010 no se dieron cursis de vacaciones y el número de cursos especiales fue de 955, con un promedio de 3.439.

En la tabla 72 se muestran las cancelaciones del periodo B de 2010 y el periodo A de 2011 de la facultad de artes, clasificadas por tipo de materia. Se mira un registro total de 713 materias canceladas. Se puede observar que el mayor registro de cancelaciones en los dos periodos es de las competencias básicas y humanísticas, 265 materias canceladas para periodo el B de 2010 y 242 materias canceladas para periodo el A de 2011. El total de materias canceladas para 2010 fue de 365 y para el A de 2011 fue de 348.

Tabla 72. Información de cancelaciones 2010/2011

Año	Tipo de materia	Materias Canceladas
	Total cancelaciones B 2010 - A 2011	713
2010	Total B 2010	365
	COMPETENCIA BÁSICA/HUMANÍSTICA	265
	NORMAL/ELECTIVA	100
2011	Total A 2011	348
	COMPETENCIA BÁSICA/HUMANÍSTICA	242
	NORMAL/ELECTIVA	106

Fuente. Esta investigación.

4.5.4 Información de profesores. A continuación se describe la información que permite analizar los datos de número de profesores en el periodo A del año 2011, esta información se detalla por cada departamento a los que pertenecen los docentes.

Tabla 73. Información de profesores por departamento

Departamento	Nivel Académico	Número de Profesores
Total	Total Nivel Académico	946
ADMINISTRACION DE EMPRESAS	Total Nivel Académico	70
ARQUITECTURA	Total Nivel Académico	29
CIENCIAS JURIDICAS	Total Nivel Académico	35
DISEÑO	Total Nivel Académico	50
ECONOMIA	Total Nivel Académico	60
ELECTRONICA	Total Nivel Académico	46
HUMANIDADES Y FILOSOFIA	Total Nivel Académico	33
LINGUISTICA E IDIOMAS	Total Nivel Académico	42
MATEMATICAS Y ESTADISTICA	Total Nivel Académico	118
MEDICINA	Total Nivel Académico	102
PROCESOS INDUSTRIALES	Total Nivel Académico	58
PRODUCCION Y PROCESAMIENTO ANIMAL	Total Nivel Académico	43
PRODUCCION Y SANIDAD VEGETAL	Total Nivel Académico	53
PROMOCION DE LA SALUD	Total Nivel Académico	34
PSICOLOGIA	Total Nivel Académico	47

Tabla 73. Continuación.

RECURSOS NATURALES Y SISTEMAS AGROFORESTALES	Total Nivel Académico	45
SALUD ANIMAL	Total Nivel Académico	35
SISTEMAS	Total Nivel Académico	46
INGENIERIA DE SISTEMAS	Total Nivel Académico	46
INGENIERIA DE SISTEMAS	DOCTOR	1
INGENIERIA DE SISTEMAS	ESPECIALISTA	24
INGENIERIA DE SISTEMAS	MAGISTER	7
INGENIERIA DE SISTEMAS	PROFESIONAL	14

Fuente. Esta investigación.

En la tabla 73 se puede analizar que en el periodo A del año 2011 el departamento con número de docentes adscritos en el departamento de Matemáticas y Estadística con 118 docentes que representan el 12% del total mientras que Arquitectura con 29 profesores es el departamento con menos profesores adscritos. También se puede analizar que en el departamento de sistemas la mayoría de sus docentes son especialistas y que solo se cuenta con un profesor con el título de doctor.

A continuación se describe la información que permite analizar los datos de número de profesores en el periodo A del año 2011, esta información se detalla por tipo de vinculación del docente con la Universidad de Nariño.

Tabla 74. Información de vinculaciones de profesores.

Vinculación	Número de Profesores
Total	946
HORA CATEDRA	275
HORA CATEDRA HONORARIOS	34
HORA CATEDRA JUBILADO	27
HORA CATEDRA POR SERVICIOS PRESTADOS	380
INVITADO	1
TIEMPO COMPLETO	185
TIEMPO COMPLETO OCASIONAL	44

Fuente. Esta investigación.

En la tabla 74 se puede analizar que en año 2011 se encuentran 946 Docentes en la Universidad de Nariño, de los cuales la mayoría están vinculados por hora cátedra, mientras se observa que solo un docente esta en calidad de invitado.

A continuación se describe la información que permite analizar las calificaciones de los profesores en la evaluación docente del año 2010, los datos están detallados por el formulario diligenciado por los estudiantes. Además por cada ítem que conforma la característica “fomento del espíritu investigativo”.

Tabla 75. Información de evaluaciones

Característica	Gran Fortaleza	Fortaleza	Transición	Debilidad	Gran Debilidad
Total	16,887	1,672	345	62	9
CONOCIMIENTOS	2,875	341	68	15	1
CONTRIBUCIÓN AL CLIMA ORGANIZACIONAL	3,059	197	39	4	1
CUMPLIMIENTO Y USO EFICIENTE DEL TIEMPO	2,899	307	78	12	4
EVALUACIÓN APLICADA A LOS ESTUDIANTES	2,911	302	73	12	2
FOMENTO DEL ESPIRITU INVESTIGATIVO	2,958	291	41	10	0
Fomenta el análisis de los problemas de la asignatura con el concurso de diferentes disciplinas	764	54	5	2	0
Fomenta el trabajo en grupo	720	91	12	2	0
Incentiva y asesora al estudiante en el análisis crítico de diversas fuentes de información(artículos, textos, internet)	725	86	10	4	0

Tabla 75. Continuación

Tiene en cuenta y respeta los saberes y opiniones de los estudiantes	749	60	14	2	0
METODOLOGIA	2,185	234	46	9	1

Fuente. Esta investigación.

En la tabla 75 se puede analizar que en los diferentes ítems de la característica FOMENTO DEL ESPIRITU INVESTIGATIVO, cerca del 90% de los docentes sacan la mejor calificación posible “Gran Fortaleza” mientras que ningún docente saca la calificación negativa “Gran Debilidad”. Con lo que se puede concluir que esta característica es un punto fuerte en los docentes de la Universidad de Nariño.

5. CONCLUSIONES

1. Como resultado de este proyecto, la Universidad de Nariño cuenta con el diseño de un *Datamart* de información académica, que una vez implementado, le permitirá contar con información académica, histórica y fidedigna. Combinando esta tecnología con herramientas de análisis multidimensional, los administrativos pueden descubrir estrategias que mejoren la toma de decisiones en el área académica de la Universidad de Nariño.
2. El diseño del *Datamart* ha permitido tener una mejor visión de los sucesos en el ámbito universitario así como dar apoyo a la toma de decisiones. En un mundo competitivo como el actual, incluso dentro del ámbito universitario, no es posible quedarse al margen de la tecnología para apoyar la administración de negocios. En este sentido, la construcción del *Datamart* juega un papel importante en la toma de decisiones.
3. El proyecto establece un marco teórico que permite conocer los fundamentos que sostienen la tecnología de *Data Warehouse*. Partiendo de las más importantes metodologías que se ofrecen actualmente, se seleccionó la metodología que mejor se adaptaba al proyecto y se ha desarrollado punto por punto para alcanzar los objetivos marcados. Una vez dentro del marco de trabajo que permitió un buen diseño, se aplicó la metodología de Ralph Kimball al proyecto *Datamart* de información académica para la Universidad de Nariño.
4. El prototipo construido a partir del diseño del *Datamart* académico y analizado mediante la herramienta *STARCUBE*, permitió comprobar y validar el diseño general del proyecto. Mediante la herramienta *STARCUBE* se pueda obtener información de este *Datamart*, importante para dar respuesta a los diferentes requerimientos de información académica de la Universidad de Nariño y mejorar la toma de decisiones por parte de los directivos.
5. Este proyecto fue dado a conocer en una ponencia en el evento “Primer Encuentro de Ponencias de Grupos de Investigación” de la Corporación Universitaria autónoma de Nariño.
6. Durante el desarrollo de la investigación se encontraron dos bases de datos transaccionales, en los SGBD *PostgreSQL* e *Informix*, con información en

ocasiones inconsistente e incompleta. Para consolidar los datos en el prototipo funcional se requirió de varios scripts en lenguaje *SQL* para migrar y limpiar los datos, además se utilizó la herramienta *EXDACLET*, desarrollada en el grupo de investigación *GRIAS*, siendo de esta manera muy complejo el proceso de carga de los datos en el prototipo.

6. RECOMENDACIONES

1. Implementar el diseño propuesto y replicar esta experiencia en la construcción de un Data Warehouse de la Universidad de Nariño, compuesto por varios datamarts que proporcione información histórica y resumida de las diferentes áreas de la Universidad que dote a sus directivas de una herramienta eficaz para soportar la toma de decisiones.
2. Construir una nueva herramienta de análisis multidimensional que se ajuste a los requerimientos propios de la Universidad de Nariño y que agilice las diferentes consultas en el área académica, desarrollada para la plataforma web, la cual hoy en día es la más usada.
3. Construir un nuevo sistema académico transaccional unificado que sirva de fuente fiable de datos para el *Datamart*.
4. Ubicar el *Datamart* en uno de los servidores de bases de datos de la Universidad de Nariño, estableciendo los parámetros de seguridad y permitiendo a los directivos acceder a estos recursos de manera remota y así soportar el proceso de la toma de decisiones.

7. BIBLIOGRAFIA

1. Avila, H.: Creación de una Bodega de Datos con información génica sobre el proceso de transcripción de la bacteria E. Coli, Instituto Tecnológico y de Estudios Superiores de Monterrey, Cuernavaca, México, (2002).
2. Bahamón, J.: Propuesta de un Método para el Diseño y Modelado de una Bodega de Datos, Universidad Icesi, Cali, Colombia, (2003).
3. Elmasri, R. y Navathe, S.: Fundamentos de Sistemas de Bases de Datos, Editorial Pearson, 5 edición, Madrid, España, (2007).
4. García, B.: Sistema de Información Gerencial Académico – Financiero para F.I.S.I.C.C (Data Mart), Universidad Francisco Marroquín, Ciudad de Guatemala, Guatemala, (2002).
5. Hernández, J. y Ramírez, M. y Ramírez, C.: Introducción a la Minería de Datos, Editorial Pearson, Madrid, España, (2004).
6. Inmon, W.: Building the Data Warehouse, Editorial Wiley Computer, New York 2 edición, (1996).
7. Mannino, M.: Administración de Bases de datos Diseño y Desarrollo de Aplicaciones, Editorial McGraw Hill, 3 edición, México, (2007).
8. Rivadera, G.: La Metodología de Kimball para el Diseño de Almacenes de Datos (*Data Warehouses*), Universidad Católica de Salta, Salta, Argentina, (2010).
9. Rodríguez, M.: Análisis Y Diseño de un Data Mart para el Seguimiento Académico de Alumnos en un Entorno Universitario, Universidad Carlos III de Madrid, Madrid, España, (2010).
10. Sevilla, E.: Guía Metodológica para la Definición y Desarrollo de un Data Warehouse, Universidad Americana, Managua, Nicaragua, (2003).
11. Sinergia e Inteligencia de Negocios. Bases de datos OLTP y OLAP. [En línea][Citado el: 15 de mayo de 2011] http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx.
12. Timarán, R.: Bodegas de Datos y Procesamiento Analítico en línea, en Electiva II Base de datos II, Universidad de Nariño, Pasto, Colombia, (2009).

13. Vizuite, W. y Yela, S.: Análisis, Diseño e Implementación de una Datamart para el Área de Sismología del Departamento de Geofísico de la Escuela Politécnica Nacional, Escuela Politécnica Nacional, Quito, Ecuador, (2006).

14. Zvenger, P.: Tesis de licenciatura: Introducción al Soporte de Decisiones: Incorporación de Soluciones Olap en Entornos Empresariales, Universidad Nacional del Sur, Argentina, (2005).

ANEXOS

ANEXO A

MARCAS REGISTRADAS UTILIZADAS.

- ProcesadorIntel®Core™ 2 Quad, Copyright ©Intel Corporation.
- ProcesadorIntel®Core™ i7, Copyright ©Intel Corporation.
- Impresora LASER Hewlett Packard™, Copyright © 2008 Hewlett-Packard Development Company, L.P.
- PostgreSQL™ v. 8.4.4, Copyright © 1996 – 2009 PostgreSQL Global Development Group.
- Microsoft® Office Excel © 2010.

ANEXO B

DESCRIPCION Y EJECUCIÓN SCRIPTS DATAMART UDENAR

1. Diccionario de datos de Información académica. Para el desarrollo del proyecto se hizo necesario el análisis de las bases de datos, teniendo en cuenta las tablas que tienen información de los procesos de inscripciones y admisiones, matrículas y población estudiantil, materias y rendimiento académico y por último el proceso de profesores.

El diccionario de datos se encuentra en la carpeta diccionario de datos, del DVD anexo a este documento, cada fuente de datos se describe en una subcarpeta con el nombre de cada uno de los procesos identificados.

2. Scripts Datamartudenar. Los scripts están definidos principalmente por los esquemas académico y profesores, los cuales contienen las fuentes de datos importantes para la creación de las dimensiones del Datamart. Los scripts se encuentran en el DVD anexo, en la carpeta Scripts, los pasos a seguir son los siguientes:

- Desde la consola de Windows dirigirse a la ruta de instalación de postgres y ejecutar el comando:

```
Createdb -U "roldesuperusuario" datamartudenar.
```

Con lo cual se creará la base de datos datamartudenar, con la codificación utf8.

- Ejecutar los comandos:

```
psql -U "roldesuperusuario" -d datamartudenar -f "rutadelscript/academico"  
psql -U "roldesuperusuario" -d datamartudenar -f "rutadelscript/profesores"
```

Con esto se habrá creado los esquemas: académico y de profesores, los cuales contienen las tablas fuente que alimentan el *Datamart*.

Para la creación de las dimensiones y tablas de hechos se necesitan de scripts, los cuales se ejecutan bajo la orden de consola "i rutascript" bajo el rol del superusuariopostgres.

- **Creación de las dimensiones comunes dentro del datamart.** Estas son las primeras dimensiones que se debe crear. La ubicación de los scripts se encuentra en la carpeta DatamartUdenar y dentro de esta la carpeta Dimensiones comunes. Se debe ejecutar cada una de los scripts temporales ubicados dentro de la carpeta scripts temporales. Por ejemplo el script tmp_alumno. Luego de ejecutar cada uno de estos scripts se procede a crear las dimensiones mediante los scripts que se encuentran en la carpeta Dimensiones. Para la dimensión alumno es importante que después de ejecutar todo el proceso de creación del *Datamart* de información académica, se ejecute el script borrar campos dentro de la carpeta Dimensiones de la parte de Dimensiones comunes.

- **Creación Datamart Inscripciones y admisiones.** Después de crear las dimensiones comunes se crean las dimensiones de inscripciones y admisiones. Para esto se debe ejecutar el script “tmp_inscrito.sql” dentro de la carpeta Scripts temporales encontrado en la carpeta “DatamartUdenar/Datamart Inscripciones y admisiones”. Luego se crea las dimensiones ejecutando cada uno de los scripts dentro de la carpeta “DatamartUdenar/Datamart Inscripciones y admisiones”. Al final se crea la tabla de hechos con la ejecución del script temporal “tmp_fact_cobertura.sql” el cual se encuentra en la carpeta de scripts temporales. Luego se crea la dimensión ejecutando el script “fact_cobertura.sql” que se encuentra en la carpeta Tablas de hechos.

- **Creación Datamart Matrículas y población estudiantil.** Se debe ejecutar el script temporal “tmp_matriculados.sql” el cual se encuentra en la carpeta “DatamartUdenar/Datamart matrículas y población estudiantil”. Luego se crea las dimensiones ejecutando los scripts que se encuentran en la ruta “DatamartUdenar/Datamart matrículas y población estudiantil/Dimensiones”. Por último se crea la tabla de hechos con el script “fact_matriculados”, el cual se encuentra en la carpeta Tablas de hechos.

- **Creación Datamart Materias y rendimiento académico.** Se debe ejecutar los scripts dentro de la carpeta Dimensiones en la carpeta “/DatamartUdenar/Datamart Materias y rendimiento académico”. Luego se crea las tablas de hechos mediante los scripts que se encuentran en la carpeta tablas de hechos, los cuales son: fact_cancelacion, fact_eventos, fact_eventos_mat, fact_rendacad.

- **Creación Datamart Profesores.** Se deben ejecutar todos los scripts temporales dentro de la ruta “DatamartUdenar/Datamart Profesores/Scripts

Temporales”. Luego se crean las dimensiones con los scripts que están dentro de la carpeta “DatamartUdenar/Datamart Profesores/Dimensiones”. Y por último se crean las tablas de hechos con los scripts de la carpeta “DatamartUdenar/Datamart Profesores/Tablas de hechos”.

El paso final es ejecutar el script borrarcampos.sql dentro de la ruta “DatamartUdenar/Dimensiones comunes/Dimensiones” el cual ya se mencionó anteriormente.

Nota: Si desea crear el repositorio con todas las dimensiones y fuentes, sin seguir los pasos nombrados anteriormente, debe ejecutar el script “datamartudenar.sql” dentro de la carpeta “Scripts”. El programa STARCUBE, para análisis multidimensional se encuentra en la carpeta StarCube del DVD anexo, dentro de esta carpeta también se encuentra el backupStarCube.backup, importante para el funcionamiento de *STARCUBE*.