

**UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL
DE HUAMANGA**

FACULTAD DE INGENIERÍA DE MINAS, GEOLOGÍA Y CIVIL

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



TESIS:

**Datamart para el seguimiento académico de los estudiantes en
la Universidad Nacional de San Cristóbal de Huamanga,
Ayacucho, 2024**

Para optar el título profesional de:
INGENIERO DE SISTEMAS

PRESENTADO POR:

Bach. Fernando ANAYA POMA

ASESOR:

Mg. Ing. Hubner JANAMPA PATILLA

AYACUCHO - PERÚ

2025

Dedicatoria

A mis padres por ser mi motivo más importante para la culminación de esta etapa y demostrarme siempre su apoyo incondicional y también formarme con buenos valores y hábitos. A mis maestros por guiarme y orientarme en mi vida estudiantil. A mis amigos que en muchas ocasiones me animaron a seguir adelante en mis proyectos.

Agradecimiento

A mis papas y hermanos que son mi fuerza e inspiración en cada uno de mis proyectos y me impulsan a culminar mis metas.

A la Universidad Nacional de San Cristóbal de Huamanga, por acogerme en sus aulas y permitir culminar mis estudios de pregrado.

Al Mg. Ing. Hubner Janampa Patilla, por apoyarme e impulsarme en esta etapa para culminar mi formación.

Resumen

En la Universidad Nacional de San Cristóbal de Huamanga, existe una cantidad considerable de datos en sus bases académicas que no se están utilizando de manera eficiente. La información relacionada con calificaciones, matrículas y desempeño académico no está siendo aprovechada de forma óptima, lo que dificulta el seguimiento adecuado del progreso estudiantil y la implementación de decisiones estratégicas. Esta falta de aprovechamiento de los datos impide que las autoridades universitarias desarrollen políticas o acciones que mejoren el rendimiento académico y brinden un mejor apoyo a los estudiantes.

Para abordar esta situación, se diseñó e implementó un Datamart. Esta herramienta permitirá centralizar, analizar y aprovechar la información académica de manera más eficiente. El objetivo principal es optimizar el seguimiento académico en la universidad, mejorando aspectos como la eficiencia en el seguimiento, la precisión en la información y la accesibilidad de los datos. De este modo, se facilitará la generación de indicadores útiles para que las autoridades puedan tomar decisiones mejor fundamentadas, lo que contribuirá a la mejora de la calidad educativa.

Este estudio presenta la implementación de un Datamart que, mediante un diseño adecuado y la integración de diversas fuentes de información, proporcionará reportes y análisis académicos de manera rápida, precisa y accesible. Como resultado, la universidad estará en condiciones de desarrollar estrategias más efectivas para mejorar el rendimiento académico de sus estudiantes. La investigación tiene como finalidad resolver los problemas actuales en la gestión y análisis de los datos académicos de la Universidad Nacional de San Cristóbal de Huamanga, proponiendo el Datamart como una herramienta clave para optimizar la toma de decisiones y mejorar la experiencia académica de los estudiantes.

Palabras clave: Datamart, seguimiento académico, Universidad Nacional de San Cristóbal de Huamanga, metodología Kimball.

Abstract

At the National University of San Cristóbal de Huamanga, there is a considerable amount of data in its academic databases that is not being used efficiently. Information related to grades, enrollment, and academic performance is not being used optimally, which makes it difficult to properly monitor student progress and implement strategic decisions. This lack of data utilization prevents university authorities from developing policies or actions that improve academic performance and provide better support to students.

To address this situation, a Datamart was designed and implemented. This tool will allow centralizing, analyzing, and using academic information more efficiently. The main objective is to optimize academic monitoring at the university, improving aspects such as monitoring efficiency, information accuracy, and data accessibility. In this way, the generation of useful indicators will be facilitated so that authorities can make better-informed decisions, which will contribute to improving educational quality.

This study presents the implementation of a Datamart that, through an appropriate design and the integration of various sources of information, will provide academic reports and analysis in a fast, accurate and accessible manner. As a result, the university will be in a position to develop more effective strategies to improve the academic performance of its students. The purpose of the research is to solve the current problems in the management and analysis of academic data at the National University of San Cristóbal de Huamanga, proposing the Datamart as a key tool to optimize decision making and improve the academic experience of students.

Keywords: Datamart, academic monitoring, National University of San Cristóbal de Huamanga, Kimball methodology.

Índice		
Resumen		v
Abstract		vi
	Capítulo I	
	Introducción	
1.1	Planteamiento del problema	15
1.2	Formulación del problema	15
1.2.1	Problema principal	15
1.2.2	Problemas secundarios	15
1.3.	Justificación de la investigación	16
1.4.	Limitaciones de la investigación	16
1.5.	Objetivos de la investigación	16
1.5.1.	Objetivo General	16
1.5.2.	Objetivos Específicos	17
1.6.	Hipótesis de la investigación	17
	Capítulo II	
	Marco teórico	
2.1.	Antecedentes de la investigación	18
2.2.	Marco teórico	19
2.2.1.	Seguimiento académico	19
2.2.1.1	Eficiencia del seguimiento	19
2.2.1.2	Precisión En La Información	19
2.2.1.2	Accesibilidad de los datos	20
2.2.2.	Datamart	20
2.2.2.1	Diseño y arquitectura del Datamart	21
2.2.2.2	Integración de datos	21
2.2.2.3	Calidad de los datos	21

2.2.3.	Ventajas y desventajas	23
2.2.4.	Modelo lógico del Data Warehouse	23
2.5.5.	Data Warehouse VS Datamart	27
2.5.6.	Proceso de extracción, transformación y carga (ETL)	28
2.5.7.	Análisis en línea OLAP (ON-LINE ANALYTICAL PROCESSING)	30
2.5.8.	Sistema gestor de base de datos (SGBD)	33
2.5.9.	Metodología Kimball	36

Capítulo III

Material y métodos

3.1.	Tipo de investigación	39
3.2.	Nivel de investigación	39
3.3.	Diseño de la investigación	39
3.4.	Población y muestra	40
3.4.1.	Población	40
3.4.2.	Muestra	40
3.5.	Variables e indicadores	40
3.5.1.	Definición conceptual de las variables	40
3.5.2.	Definición operacional de las variables	41
3.6.	Técnicas e instrumentos para el tratamiento de datos e información	42
3.6.1.	Técnicas para recolectar información	42
3.6.2.	Instrumentos para recolectar información	42
3.6.3.	Herramientas para el tratamiento de datos e información	42
3.7.	Técnicas para aplicar la metodología Kimball	43

Capitulo IV

Resultados de la investigación

4.1.	Resultados aplicando la metodología Kimball	47
4.1.1.	Planificación del proyecto	47

4.1.2.	Definición de los requerimientos	47
4.1.2.1.	Requerimientos funcionales	47
4.1.2.2.	Requerimientos no funcionales	48
4.1.3.	Diseño de la arquitectura técnica	49
4.1.4.	Selección del producto	50
4.1.5.	Modelo dimensional	50
4.1.5.1.	Desarrollo del modelo dimensional	50
4.1.5.2.	Diagramas de las tablas de dimensiones	54
4.1.5.3.	Diagrama de la tabla de hecho	60
4.1.5.4.	Diagrama del modelo lógico	62
4.1.6.	Diseño del modelo físico	65
4.1.6.1.	Construcción de las tablas y la base de datos	69
4.1.7.	Diseño e implementación del sub sistema ETL	73
4.1.7.1.	Fases del proceso ETL	73
4.1.7.2.	Pasos del proceso ETL	75
4.1.8.	Implementación del cubo multidimensional	97
4.1.9.	Especificación de aplicación para los usuarios finales	105
4.1.10.	Desarrollo de la aplicación en el Power Bi	106
4.1.11.	Crecimiento y mantenimiento del Datamart	113

Capitulo V

Conclusiones y recomendaciones

5.1.	Conclusiones	114
5.1.	Recomendaciones	114
	Referencias bibliográficas	115
	Anexo 1: Operacionalización de variables	119
	Anexo 2: Fichas de análisis para la BD de la universidad	121
	Anexo 3: Guía de entrevista para el Datamart	123

Lista de tablas

Tabla 1 Instrumentos tipos de aplicaciones utilizadas para el manejo de los datos	42
Tabla 2 Programación del proyecto del Datamart	43
Tabla 3 Determinar los requerimientos	43
Tabla 4 Diseño de la arquitectura técnica	44
Tabla 5 Selección de producto	44
Tabla 6 Modelo dimensional	44
Tabla 7 Diseño físico	45
Tabla 8 Proceso ETL	45
Tabla 9 Implementación	46
Tabla 10 Especificación de aplicaciones para usuarios finales	46
Tabla 11 Mantenimiento y crecimiento	46
Tabla 12 Lista de requerimientos	47
Tabla 13 Software para la construcción del Datamart	50
Tabla 14 Lista de indicadores	51
Tabla 15 Lista de perspectivas	51
Tabla 16 Objetivo de la tabla de hechos	53
Tabla 17 Dimensiones de la tabla de hechos	53
Tabla 18 Elección de hechos	54
Tabla 19 Dimensión Facultad/Escuela	55
Tabla 20 Dimensión Estudiante	56
Tabla 21 Dimensión Curso	57
Tabla 22 Dimensión Periodo académico	58
Tabla 23 Dimensión Plan Estudio	58
Tabla 24 Dimensión Matricula	59
Tabla 25 Detalle de las llaves de las dimensiones	61
Tabla 26 Detalles de las medidas de la tabla de hecho Fact_SeguimientoAcademico	62

Tabla 27	Tipo de datos para Dim_Estudiente	65
Tabla 28	Tipo de datos para Dim_EscuelaFacultad	65
Tabla 29	Tipo de datos para Dim_Curso	66
Tabla 30	Tipo de datos para Dim_PeriodoAcademico	66
Tabla 31	Tipo de datos para Dim_PlanEstudio	66
Tabla 32	Tipo de datos para Dim_Matricula	67

Lista de figuras

Figura 1 <i>Datamart dependiente</i>	21
Figura 2 <i>Datamart independiente</i>	22
Figura 3 <i>Tablas de dimensiones</i>	23
Figura 4 <i>Tablas de hechos</i>	24
Figura 5 <i>Uniones de tablas</i>	24
Figura 6 <i>Esquema en estrella</i>	25
Figura 7 <i>Variante de esquema estrella</i>	27
Figura 8 <i>Esquema copo de nieve</i>	27
Figura 9 <i>Diferencias entre Data Warehouse VS Datamart</i>	28
Figura 10 <i>Caso práctico, carga Inicial</i>	28
Figura 11 <i>Diferencias entre MOLAP, ROLAP, HOLAP</i>	33
Figura 12 <i>Ciclo de vida Kimball</i>	37
Figura 13 <i>Arquitectura técnica</i>	49
Figura 14 <i>Diagrama del modelo conceptual inicial</i>	52
Figura 15 <i>Jerarquía de atributos de la dimensión EscuelaFacultad</i>	54
Figura 16 <i>Jerarquía de atributos de la dimensión Estudiante</i>	55
Figura 17 <i>Jerarquía de atributos de la dimensión Curso</i>	56
Figura 18 <i>Jerarquía de atributos de la dimensión Periodo Académico</i>	57
Figura 19 <i>Jerarquía de atributos de la dimensión Plan Estudio</i>	58
Figura 20 <i>Jerarquía de atributos de la dimensión Matricula</i>	59
Figura 21 <i>Diagrama de la tabla de hechos</i>	60
Figura 22 <i>Tabla de hecho del Datamart</i>	62
Figura 23 <i>Modelo lógico de Datamart</i>	64
Figura 24 <i>Diseño del modelo físico del Datamart</i>	68
Figura 25 <i>Tabla Dim_Estudiante</i>	69
Figura 26 <i>Tabla Dim_EscuelaFacultad</i>	69

Figura 27	<i>Tabla Dim_Curso</i>	69
Figura 28	<i>Tabla Dim_PeriodoAcademico</i>	70
Figura 29	<i>Tabla Dim_PlanEstudio</i>	70
Figura 30	<i>Tabla Dim_Matricula</i>	70
Figura 31	<i>Tabla Fact_SeguimientoAcademico</i>	71
Figura 32	<i>Modelo físico del Datamart en SQL Server</i>	72
Figura 33	<i>Diagrama de pasos del proceso ETL</i>	74
Figura 34	<i>Diagrama de pasos con restricciones de precedencia</i>	75
Figura 35	<i>Proceso ETL de la carga inicial</i>	76
Figura 36	<i>Comando SQL para la limpieza de las tablas del Datamart</i>	77
Figura 37	<i>Origen de datos para la dimensión Estudiante</i>	77
Figura 38	<i>Conversión de datos para la dimensión Estudiante</i>	78
Figura 39	<i>Carga de datos para la dimensión Estudiante</i>	79
Figura 40	<i>Origen de datos para la dimensión Escuela Facultad</i>	80
Figura 41	<i>Conversión de datos para la dimensión Escuela Facultad</i>	81
Figura 42	<i>Carga de datos para la dimensión Escuela Facultad</i>	82
Figura 43	<i>Origen de datos para la dimensión Curso</i>	83
Figura 44	<i>Conversión de datos para la dimensión Curso</i>	84
Figura 45	<i>Carga de datos para la dimensión Curso</i>	85
Figura 46	<i>Origen de datos para la dimensión Periodo Académico</i>	86
Figura 47	<i>Conversión de datos para la dimensión Periodo Académico</i>	87
Figura 48	<i>Carga de datos para la dimensión Periodo Académico</i>	88
Figura 49	<i>Origen de datos para la dimensión Plan Estudio</i>	89
Figura 50	<i>Conversión de datos para la dimensión Plan Estudio</i>	90
Figura 51	<i>Carga de datos para la dimensión Plan Estudio</i>	91
Figura 52	<i>Origen de datos para la dimensión Matricula</i>	92
Figura 53	<i>Conversión de datos para la dimensión Matricula</i>	93

Figura 54	<i>Carga de datos para la dimensión Matricula</i>	94
Figura 55	<i>Proceso ETL completado para el Datamart</i>	96
Figura 56	<i>Administrador de la conexión con el Datamart</i>	97
Figura 57	<i>Selección de tablas para la generación de las vistas</i>	98
Figura 58	<i>Dimensiones y tabla de hechos para generar la vista</i>	99
Figura 59	<i>Vista de origen del esquema estrella del Datamart</i>	100
Figura 60	<i>Selección de la tabla principal</i>	101
Figura 61	<i>Tabla principal con sus medidas</i>	102
Figura 62	<i>Creación del cubo dimensional y la relación con sus dimensiones</i>	103
Figura 63	<i>Diagrama final del cubo de Seguimiento académico</i>	104
Figura 64	<i>Usuarios asignados al Datamart de seguimiento académico</i>	105
Figura 65	<i>Configuración de la conexión con el Power Bi</i>	106
Figura 66	<i>Configuración del servidor para la conexión con Power Bi</i>	107
Figura 67	<i>Cantidad de matriculados por género, escuela y facultad</i>	108
Figura 68	<i>Cantidad de veces que llevo un alumno sus cursos</i>	109
Figura 69	<i>Cantidad de aprobados y desaprobados por curso, escuela y periodo</i>	110
Figura 70	<i>Notas por curso y periodo académico del estudiante</i>	111
Figura 71	<i>Tasa de éxito y fracaso de un curso por periodo</i>	112

Capítulo I

Introducción

1.1 Planteamiento del problema

Actualmente, la Universidad Nacional de San Cristóbal De Huamanga enfrenta dificultades para explotar eficazmente la información disponible en sus bases de datos académicas. Esta información, que incluye calificaciones, cantidad de estudiantes matriculados, asignaturas, cursos aprobados y desaprobados, no está siendo aprovechada de manera óptima para realizar un seguimiento adecuado del rendimiento académico de los estudiantes. La falta de una herramienta eficiente para gestionar y analizar estos datos limita la capacidad de las autoridades universitarias para tomar decisiones estratégicas que puedan mejorar el rendimiento y el éxito académico de los estudiantes.

La ausencia de un Datamart que permita analizar y explotar de manera eficiente la información académica existente, afecta la capacidad de la universidad para generar indicadores de rendimiento académico. Esto impide que las autoridades puedan tomar decisiones informadas para mejorar la calidad educativa, optimizar el rendimiento estudiantil y desarrollar estrategias efectivas de apoyo a los estudiantes. Dado el gran volumen de datos generados en los procesos académicos, es esencial implementar una herramienta que centralice, procese y analice la información para obtener indicadores clave que faciliten la mejora continua del proceso educativo.

1.2 Formulación del problema

1.2.1 Problema principal

¿Cómo se puede diseñar e implementar un Datamart que realice el seguimiento académico de los estudiantes en la Universidad Nacional De San Cristóbal De Huamanga, Ayacucho, 2024?

1.2.2 Problemas secundarios

- a. ¿Cómo el Datamart mejorará la **eficiencia del seguimiento** para la toma de decisiones en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024?

- b. ¿Cómo el Datamart asegurará la **precisión en la información** del seguimiento para la toma de decisiones en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024?
- c. ¿Cómo el Datamart garantizará la **accesibilidad de los datos** académicos para la toma de decisiones en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024?

1.3. Justificación de la investigación

En esta era de cambios tecnológicos es importante usar las herramientas disponibles para poder obtener datos y procesarlos para tener información relevante y que nos proporcione un marco de apoyo para diseñar estrategias y aplicar las correcciones necesarias en los ámbitos necesarios. Para la Universidad Nacional de San Cristóbal de Huamanga, se busca diseñar e implementar un Datamart de seguimiento académico que permita usar la base de datos actual y procesar la información para que las autoridades pertinentes hagan una evaluación del ámbito académico, evaluar nuevas estrategias para mitigar los impases que presentaran los estudiantes en su desempeño estudiantil.

1.4. Limitaciones de la investigación

En la investigación se abarco el rendimiento estudiantil de las diferentes escuelas profesionales, cabe señalar que la base de datos proporcionada nos muestra un panorama del comportamiento del rendimiento, pero falta hacer un estudio más exhaustivo de otros factores que también afectan el rendimiento estudiantil, además que hay registros incompletos ya que la modernización de la universidad en cuanto al procesamiento de datos es aún inicial, eso indica que muchos registros de los estudiantes aún se encuentran en registros físico, los cuales se encuentran dispersos en las diferentes dependencias y escuelas que archivan las actas físicas, lo cual ha dificultado un análisis que abarque más periodos académicos, ya que no están completos los registros.

1.5. Objetivos de la investigación

1.5.1. Objetivo General

Diseñar e implementar un Datamart que mejore el seguimiento académico de los estudiantes en la Universidad Nacional De San Cristóbal De Huamanga, Ayacucho,

optimizando la eficiencia del seguimiento, la precisión de la información y la accesibilidad de los datos para la toma de decisiones académicas 2024.

1.5.2. Objetivos Específicos

- a. Implementar un Datamart que mejore la **eficiencia del seguimiento** en el monitoreo académico de los estudiantes contribuyendo al manejo de estrategias en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024.
- b. Implementar un Datamart que asegure la **precisión en la información** en el monitoreo académico de los estudiantes contribuyendo al manejo de estrategias en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024.
- c. Implementar un Datamart que garantice la **accesibilidad de los datos** en el monitoreo académico de los estudiantes contribuyendo al manejo de estrategias en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024.

1.6. Hipótesis de la investigación

Las investigaciones cuantitativas no necesariamente tienen una hipótesis. La decisión de formularlas o no depende fundamentalmente del objetivo principal del estudio. Las investigaciones cuantitativas que plantean hipótesis suelen ser aquellas con un enfoque correlacional o explicativo, o, en algunos casos, investigaciones descriptivas que buscan anticipar un dato o un acontecimiento (Hernández, Baptista & Collado, 201, p. 10).

Capítulo II

Marco teórico

2.1. Antecedentes de la investigación

Sarmiento (2018) plantea en su investigación sobre inteligencia de negocios aplicada a la gestión académica que su modelo proporciona información resumida que facilita el análisis de datos de manera ágil y en tiempo real. Esto permite a las autoridades universitarias tomar decisiones y realizar ajustes necesarios en el seguimiento y rendimiento académico de los estudiantes. Además, destaca que la cantidad de reportes obtenidos tras la implementación de la inteligencia de negocios supera la cantidad generada previamente, debido a que la aplicación centraliza, procesa y gestiona datos históricos sobre matrículas, desempeño académico, docentes, graduados, titulados y la población estudiantil, tanto de ingresantes como de egresados.

Por su parte, Durán (2017) señala que, en su estudio sobre el seguimiento académico de los estudiantes de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional de Cajamarca, su objetivo es apoyar la toma de decisiones educativas utilizando un Datamart, basado en la metodología de Ralph Kimball. Tras analizar la situación actual, identificó limitaciones en el sistema proporcionado por el Departamento de Tecnologías de la Información de la universidad, especialmente en cuanto a la disponibilidad de información y su acceso, lo que retrasa el análisis educativo y, en consecuencia, la toma de decisiones.

Rodríguez (2010), en su trabajo de fin de carrera, se enfoca en el análisis y diseño de un Datamart orientado al seguimiento académico en el ámbito universitario. En su investigación, plantea un marco metodológico para la implementación de proyectos de Data Warehouse y Datamart. Este enfoque, aplicado en un caso de estudio particular, permite abordar las necesidades del área académica de una universidad, optimizando la extracción de datos sobre el rendimiento de los estudiantes. Gracias a este modelo, se logra obtener una visión más clara de los acontecimientos dentro del entorno universitario, lo que ha contribuido a la toma de decisiones más acertadas en relación al desempeño académico de los estudiantes.

2.2. Marco teórico

2.2.1. Seguimiento académico

Su objetivo es mejorar el rendimiento. Los buenos avances deben complementarse con datos recopilados de encuestas, talleres de discusión o entrevistas, con los participantes del estudio (docentes, estudiantes y administradores escolares), así como expertos de otras especialidades, como profesionales en psicología, sociología, trabajo social, docencia, etc (Ruiz et. al, 2007).

El apoyo estudiantil al estudiante de pregrado es un tema importante para mejorar la calidad del servicio, y la universidad se interesa por los resultados de los estudios, la enseñanza y la evaluación de sus estudiantes que son herramientas confiables para construir señales que conlleven a tomar decisiones al respecto. enfatizar la calidad en la educación (Pérez y Samaniego, 2014).

El déficit intelectual de los estudiantes, en todos sus niveles, ya sea porque presente una asistencia irregular, o por trabajo, o por incapacidad, o por falta de docentes, o por métodos de enseñanza-aprendizaje obsoletos, o por falta de libros o por cualquier otra causa que impida el ingreso y permanencia en las calificaciones del sistema, dejando en el proceso de preparación académica de cada alumno una serie de pausas prolongadas en su avance por los niveles del sistema educativo (Reyes, 2004).

2.2.1.1 Eficiencia del seguimiento

Para Durán (2017) menciona que, mediante la implementación de cubos multidimensionales en el Datamart, la generación de consultas y reportes sobre rendimiento académico pueden realizarse de forma más ágil.

2.2.1.2 Precisión En La Información

Según Castillo, Fernández, y Martínez (2019), la exactitud en los datos se mejora mediante la implementación de algoritmos de validación y limpieza de datos, lo que logra suprimir inconsistencias y errores.

2.2.1.2 Accesibilidad de los datos

Durán (2017) señala que el diseño de un Datamart debe enfocarse en permitir que los usuarios encargados del seguimiento académico puedan acceder fácilmente a la información clave. Para ello, es fundamental contar con interfaces intuitivas y herramientas como Power BI, las cuales facilitan la creación de reportes sin que los usuarios requieran conocimientos técnicos especializados.

2.2.2. Datamart

Yalán y Palomino (2013) explican que un Datamart proporciona una base de datos dedicada a nivel departamental, diseñada para almacenar información relacionada con un área específica de la organización. Se caracteriza por proporcionar una estructura completa de datos que facilita un análisis detallado de las áreas que impactan en las operaciones del departamento. Además, un Datamart puede extraer datos de una única fuente o integrar información de varias fuentes.

Por otro lado, Silberschatz (2005) describe el Datamart como una versión específica de un repositorio de la base de datos (data warehouse). Estos subconjuntos de información están diseñados para contribuir a la formulación de estrategias en áreas concretas del negocio. Los datos pueden organizarse, explorarse y distribuirse de diversas maneras, lo que permite a los usuarios utilizarlos según sus necesidades particulares.

Moss y Atre (2003) sostienen que un Datamart es un subconjunto de un Data Warehouse, creado para satisfacer las necesidades analíticas de una unidad de negocio en particular. A diferencia de un Data Warehouse, el Datamart tiene un enfoque más específico y limitado.

Finalmente, Vitt y Luckevich (2002) definen el Datamart como una colección de datos debidamente estructurados que consume datos de diversas aplicaciones operativas. Al ser un subconjunto de un Data Warehouse, su contenido se limita a una temática o área específica dentro de la organización.

2.2.2.1 Diseño y arquitectura del Datamart

Kimball y Ross (2013) señalan que los Datamarts deben diseñarse para soportar consultas rápidas y detalladas, priorizando tanto la escalabilidad como el acceso eficiente a la información.

2.2.2.2 Integración de datos

Puryear (1993) destaca la importancia de las técnicas ETL (Extracción, Transformación y Carga) para la integración de datos, asegurando que la información sea siempre consistente y actualizada, un factor esencial para la toma de decisiones académicas informadas.

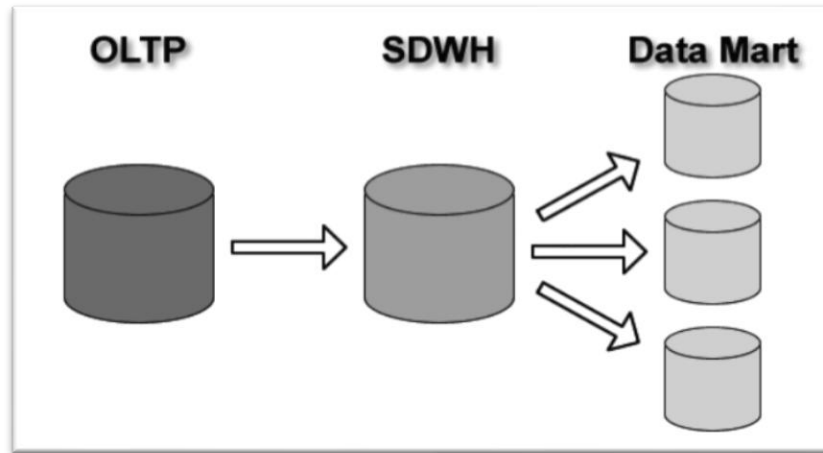
2.2.2.3 Calidad de los datos

Castillo et al. (2019) mencionan que la calidad de los datos abarca aspectos como la precisión, integridad, coherencia y puntualidad de la información. En el ámbito académico, garantizar la calidad de los datos es crucial para tomar decisiones confiables sobre el seguimiento estudiantil. La implementación de un Datamart con una adecuada gestión de la calidad de los datos puede evitar errores comunes en los informes y mejorar la toma de decisiones basada en datos.

De acuerdo con la clasificación de Inmon (2002), los Datamarts dependientes son aquellos que obtienen su información directamente de un Data Warehouse centralizado, utilizando esta única fuente de datos para su funcionamiento.

Figura 1

Datamart dependiente

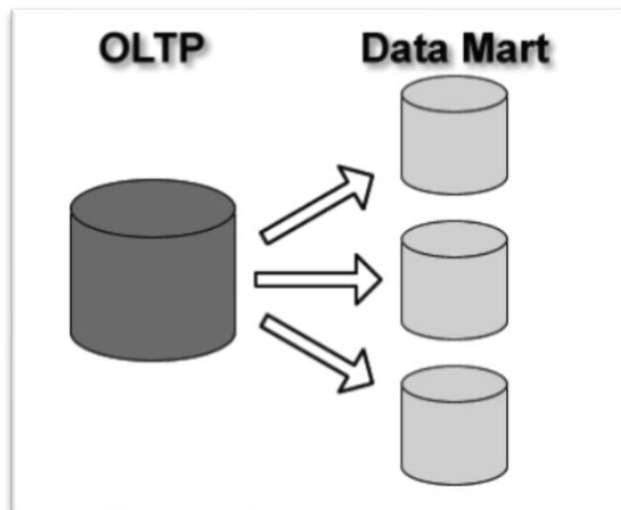


Nota: Tomada de Inmon, 2002

Los Datamarts independientes se distinguen por extraer sus datos directamente de los sistemas transaccionales, sin requerir un Data Warehouse como intermediario. Generalmente, este tipo de Datamart se nutre de la información que generan las propias organizaciones (Inmon, 2002).

Figura 2

Datamart independiente



Nota: Tomada de Inmon, 2002.

2.2.3. Ventajas y desventajas

Ventajas

De acuerdo con Nader (2002), los Datamarts ofrecen varias ventajas clave. En primer lugar, al estar dirigidos a un grupo más reducido de usuarios en comparación con un Data Warehouse, su rendimiento puede optimizarse para una recuperación de datos más eficiente. Además, al manejar un volumen menor de información, tanto la carga de datos como las consultas se procesan más rápidamente. Otra ventaja es que las solicitudes de datos pueden limitarse al área o red que gestiona esa información, lo que minimiza el impacto en otros usuarios. Asimismo, las aplicaciones cliente que realizan consultas operan de manera independiente del servidor que las procesa y del servidor de bases de datos donde se almacena la información. Finalmente, los costos de implementación de un Datamart son considerablemente más bajos que los asociados a un Data Warehouse.

Desventajas. Según Vizuite y Yela (2006), una limitación de los Datamarts es que no son ideales para manejar grandes volúmenes de información, lo que puede obligar a implementar varios Datamarts para cubrir todas las necesidades informativas de una empresa.

2.2.4. Modelo lógico del Data Warehouse

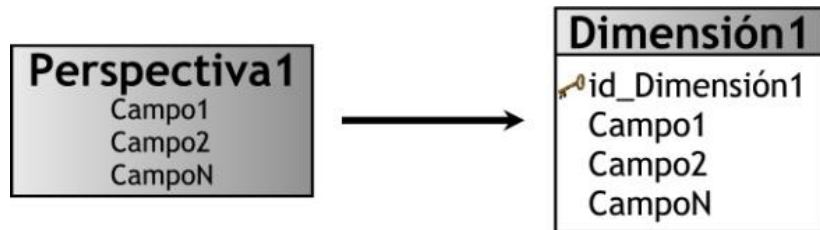
Según Bernabeu (2010), en un modelo lógico de un Data Warehouse se construye a partir de la estructura previamente definida del almacén de datos, tomando como base el modelo conceptual ya diseñado. Este proceso comienza eligiendo el tipo de esquema más adecuado y luego se realizan las acciones correspondientes para crear tanto las tablas de dimensiones como las de hechos, además de establecer las relaciones entre estas. Los principales elementos que componen el modelo lógico del Data Warehouse son: a) Tipo de modelo lógico del DW: Es necesario seleccionar el esquema más apropiado (estrella, constelación o copo de nieve), esta selección afecta de manera directa al desarrollo del modelo lógico; b) Tablas de dimensiones: Se asigna un nombre identificador a cada tabla de dimensión, se añade un campo como clave primaria, y se ajustan los nombres de los campos si es necesario para garantizar claridad; c) Tablas de hechos: En esta fase, se definen las tablas de hechos, que contienen los datos sobre los cuales se realizarán los

análisis y se generarán los indicadores; d) Relaciones: En cualquier tipo de esquema seleccionado, es fundamental establecer correctamente las uniones entre las tablas de dimensiones y las tablas de hechos para garantizar la integridad de los datos.

Figura

3

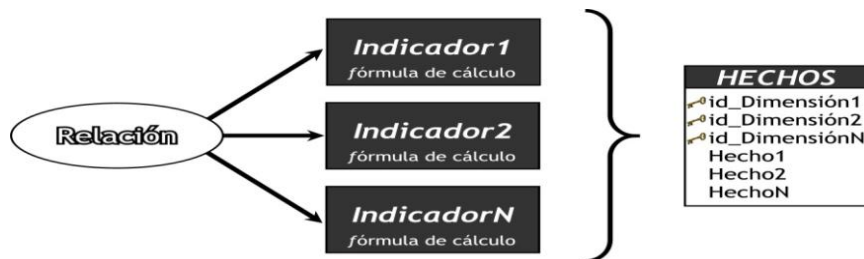
Tablas de dimensiones



Nota: Tomada de Bernabeu, 2010.

Figura 4

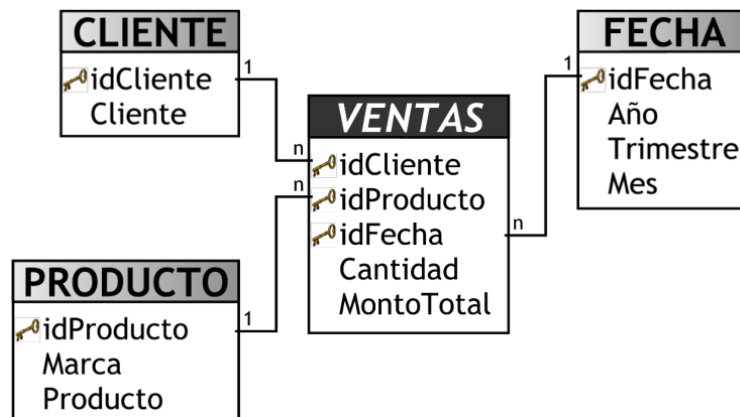
Tablas de hechos



Nota: Tomada de Bernabeu, 2010.

Figura 5

Uniones de tablas



Nota: Tomada de Bernabeu, 2010.

Inmon (2002) describe un Data Warehouse como una colección de datos que se distinguen por estar organizados en torno a temas específicos, ser integrados, no volátiles y contener registros históricos, todo ello enfocado a satisfacer las necesidades de la empresa.

Por otro lado, el INEI (2005) señala que un Data Warehouse se construye mediante la extracción de información de una o más bases de datos de sistemas transaccionales. Estos datos son transformados para eliminar inconsistencias y, en algunos casos, resumidos antes de ser cargados en el Data Warehouse. Este proceso, que incluye la transformación, la creación de un detalle temporal variable, la combinación y el resumen de los datos, facilita el acceso a la información necesaria para que las personas, a todos los niveles de la organización, puedan tomar decisiones con mayor responsabilidad.

Valdiviezo, Herrera y Jáuregui (2007) añaden que la estructura del Data Warehouse debe estar diseñada de manera que provea a los usuarios la información requerida. Es crucial realizar un análisis detallado para que los desarrolladores definan el modelo adecuado según los requerimientos que se recopilan al inicio por parte del usuario. Si la estructura del Data Warehouse no cubre todas las necesidades, los desarrolladores podrían verse obligados a volver al análisis inicial para redefinir la estructura y corregir errores.

Tipos de modelamiento de un Data Warehouse

A. Esquema estrella

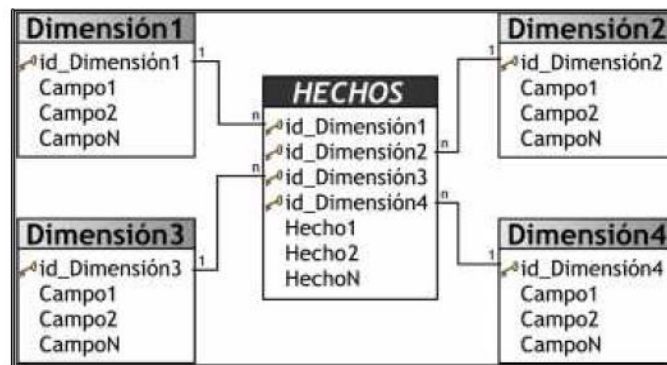
Según Ullman y Widom (1999), el esquema estrella se representa gráficamente como una estrella, donde en el centro se ubican las tablas de hechos y los conectores de la estrella están conectadas a las tablas de dimensiones.

Este modelo se distingue por tener una tabla principal, conocida como la tabla de hechos, que está rodeada por un conjunto de tablas denominadas tablas de dimensiones. En los extremos de esta estructura en forma de estrella se ubican las tablas de dimensiones, las cuales almacenan los atributos clave del negocio, permitiendo que se utilicen como filtros en las consultas. Estas tablas suelen ser de menor tamaño en comparación con la tabla de

hechos. Cada tabla de dimensión está conectada a la tabla de hechos mediante un identificador único (Kimball y Ross, 2013).

Figura 6

Esquema en estrella



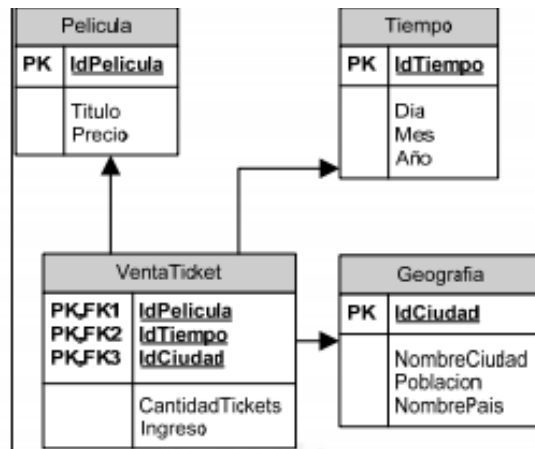
Nota: Tomada de Bernabeu, 2010.

B. Esquema copo de nieve

De acuerdo con Kimball y Ross (2013), la siguiente imagen muestra una variante del esquema estrella en la que las tablas de dimensión se encuentran normalizadas. Esto implica que las tablas pueden contener claves que se vinculan con otras tablas de dimensión. La normalización tiene ciertos beneficios, como la reducción del tamaño de las tablas y la eliminación de redundancias, además de proporcionar más flexibilidad al definir las dimensiones. No obstante, este enfoque también conlleva algunas desventajas, como la necesidad de realizar más uniones en las consultas, lo cual puede impactar negativamente en el rendimiento. Asimismo, el uso de tablas adicionales genera una mayor carga de mantenimiento.

Figura 7

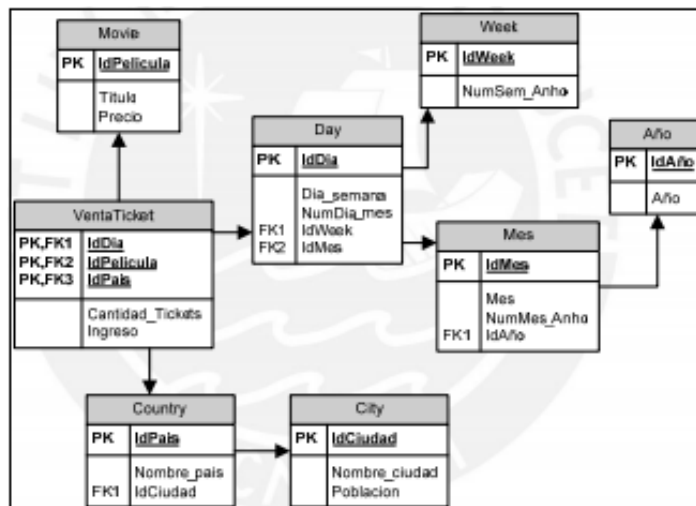
Variante de esquema estrella



Nota: Tomada de Kimball y Ross, 2013.

Figura 8

Esquema copo de nieve



Nota: Tomada de Kimball y Ross, 2013.

2.5.5. Data Warehouse VS Datamart

Los datamarts son repositorios de datos especializados, esquematizados para satisfacer las necesidades de análisis específicas de una área funcional dentro de una organización, como marketing, finanzas o producción. Estos almacenes manejan un menor número de

usuarios y cantidades de datos en comparación con un Data Warehouse centralizado, lo que les permite optimizar la carga y recuperación de información de manera más ágil y eficiente que un Data Warehouse (Moss y Atre, 2003). A continuación, se presenta una tabla que muestra las diferencias entre un Data Warehouse y un datamart.

Figura 9

Diferencias entre Data Warehouse VS Datamart

	DATA WAREHOUSE	DATAMART
Alcance	Construido para satisfacer las necesidades de información de toda la organización.	Construido para satisfacer las necesidades de un área de negocios específica.
Objetivo	Diseñado para optimizar la integración y la administración de los datos fuente.	Diseñado para optimizar la entrega de la información de soporte de decisiones.
Características de los datos	Administra grandes cantidades de datos históricos a nivel atómico.	Se concentra en administrar resúmenes y/o datos totalizados.
Pertenencia	Pertenece a toda la organización.	Pertenece al área de negocio al cual está orientado.
Administración	Es administrado por la unidad de sistema de la organización.	Es administrado por el personal de sistema de la unidad propietaria del Datamart.

Nota: Tomada de Yalán y Palomino, 2012.

2.5.6. Proceso de extracción, transformación y carga (ETL)

Según Bernabéu (2010) destaca que la consolidación de datos implica la carga de información en el sistema, utilizando técnicas de limpieza y control de calidad, junto con los procesos de extracción, transformación y carga (ETL). Posteriormente, es esencial definir las reglas y políticas necesarias para la actualización de los datos, además de los procesos encargados de llevar a cabo estas tareas.

i. Componentes de la integración de datos:

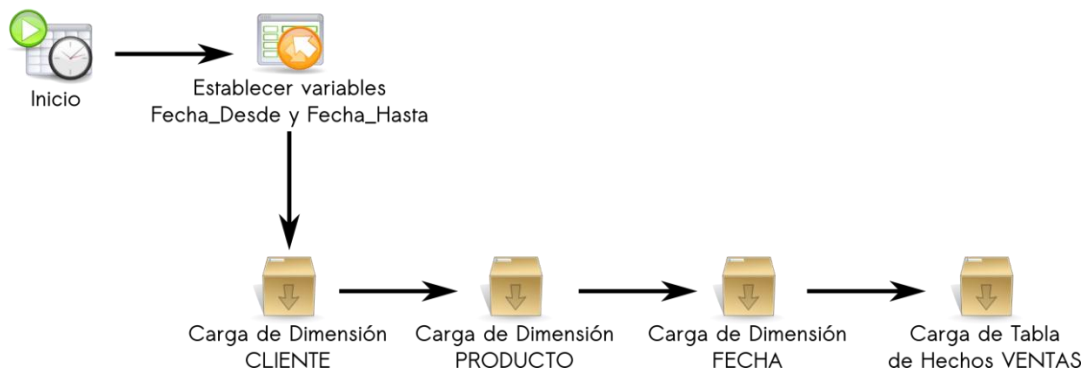
Carga inicial: Durante esta fase, se procede a cargar el datamart alimentando el modelo de datos previamente definido. Este proceso involucra tareas clave como la limpieza y

verificación de calidad de los datos, además de los procesos ETL. En algunas circunstancias, estas actividades pueden requerir una lógica compleja.

Actualización: Tras la carga inicial, es fundamental establecer estrategias y políticas para la actualización continua de los datos. Una vez definidas estas políticas, se deben implementar acciones adicionales que incluyan tareas específicas de limpieza de datos, aseguramiento de calidad y ETL para garantizar que el Data Warehouse esté actualizado. Además, es importante detallar las funciones que debe realizar cada software involucrado en el proceso.

Figura 10

Caso práctico, carga Inicial



Nota: Tomada de Bernabeu, 2010.

Zambrano (2011) indica que los procesos ETL (Extraction, Transformation, Load) son los responsables de extraer información de diversas fuentes, ya sean internas o externas, para luego transformarla mediante procesos de limpieza y consolidación, y finalmente almacenarla en un Data Warehouse o datamart.

Por otro lado, Cabanillas (2011) señala que el desarrollo de un sistema ETL consta de tres etapas fundamentales:

Extracción: Es la fase en la que se obtienen los datos necesarios para poblar el modelo físico de datos.

Transformación: Aquí se procesan los datos obtenidos para convertirlos a un formato estándar, además de calcular las métricas necesarias.

Carga: Finalmente, en esta etapa, se ejecutan los procesos para introducir los datos en los datamarts con la información previamente extraída.

Bernabéu (2010) agrega que los sistemas ETL tienen como función recopilar información de diferentes fuentes, realizar las transformaciones necesarias y almacenarla de manera integrada en un Data Warehouse. Estas transformaciones procuran salvaguardar la integridad de los datos en el repositorio. Las principales operaciones que se llevan a cabo incluyen la integración, filtrado y depuración de datos. El diseño de los procesos ETL se apoya en herramientas especializadas que facilitan la creación de sentencias SQL para extraer la información requerida de las fuentes de datos. También es posible incluir procedimientos adicionales que ayuden en el mantenimiento del modelo de datos durante los procesos ETL.

2.5.7. Análisis en línea OLAP (ON-LINE ANALYTICAL PROCESSING)

Para Laudon y Laudon (2008), OLAP muestra una perspectiva multidimensional desde un análisis de datos, lo que brinda a los usuarios la capacidad de ver la misma información desde diferentes ángulos y obtener respuestas rápidas a consultas específicas, incluso cuando trabajan con grandes volúmenes de datos. OLAP organiza las relaciones entre los datos y los cubos para realizar análisis avanzados.

Nima (2009) define OLAP como una tecnología que ofrece acceso eficiente a los datos para aplicaciones cliente.

En línea con Laudon y Laudon (2008), OLAP se enfoca en aplicaciones que analizan datos comerciales, generando información útil tanto a nivel estratégico como táctico, y apoyando así la toma de decisiones. Estas aplicaciones son flexibles y trabajan de manera óptima sobre bases de datos multidimensionales, utilizando cubos OLAP que se construyen mediante la adición de dimensiones e indicadores específicos según las necesidades de cada área o departamento.

Por su parte, Vitt y Luckevich (2002) destacan que OLAP ofrece un modelo de datos fácil de comprender, lo que permite a usuarios sin experiencia analítica interactuar con la información de manera intuitiva. Este enfoque de análisis multidimensional facilita la exploración de datos desde varias perspectivas o filtros.

Existen distintas variantes de OLAP, que se ajustan al tamaño de los datos y la eficiencia requerida. Sin embargo, OLAP no es ideal para consultas muy complejas que involucren múltiples tablas.

A. MOLAP (Multidimensional online analytical processing).

Vitt y Luckevich (2002) explican que MOLAP ofrece el mejor rendimiento en la recuperación de información, ya que organiza los datos en estructuras especializadas, alojadas en un servidor central.

Moss y Atre (2003) señalan que la ventaja de MOLAP reside en el hecho de que sus estructuras están precalculadas, lo que permite realizar las mismas consultas sobre una base estable. No obstante, como no utiliza una base de datos relacional, cualquier modificación en una parte de su estructura puede requerir la recarga de grandes cantidades de datos, lo que reduce su flexibilidad pero mejora su capacidad de respuesta.

Torres (2007) identifica varias ventajas y desventajas del MOLAP:

Ventajas: a) Rendimiento excelente: Los cubos MOLAP están optimizados para una rápida recuperación de datos. b) Capacidad para realizar cálculos complejos: Debido a que los cálculos se efectúan durante la creación del cubo, los resultados pueden obtenerse rápidamente.

Desventajas: MOLAP tiene limitaciones en cuanto al volumen de datos que puede manejar. Dado que los cálculos se generan al construir el cubo, no es posible incluir grandes cantidades de datos directamente en él. Sin embargo, los datos que se incorporan en el cubo pueden provenir de una base de datos más grande, lo que permite trabajar con un resumen de alto nivel de información.

B. ROLAP (Relational online analytical processing)

Vitt y Luckevich (2002) destacan que ROLAP ofrece la ventaja significativa de almacenar grandes volúmenes de datos. Sin embargo, su rendimiento en la recuperación de información no suele ser tan rápido como el de otras alternativas de almacenamiento, lo que lo hace más adecuado para consultas pesadas que no se realizan con frecuencia.

Tamayo y Moreno (2006) definen ROLAP como la abreviación de Relational Online Analytical Processing, que almacena datos relacionales, proporcionando así una mayor flexibilidad en los tipos de análisis disponibles y reduciendo el tiempo necesario para generar informes, además de permitir el análisis de grandes cantidades de datos. Esta tecnología se implementa en tablas físicas que están diseñadas de acuerdo con un esquema en estrella o copo de nieve.

Torres (2007) indica las ventajas y desventajas del ROLAP:

Ventajas:

Capacidad para guardar grandes volúmenes de datos: La única limitación de tamaño para la tecnología ROLAP es la que impone la base de datos relacional subyacente. Esto significa que ROLAP en sí misma no tiene restricciones de capacidad y puede aprovechar todas las funciones inherentes de la base de datos relacional.

Desventajas:

Rendimiento limitado: Como ROLAP depende de múltiples consultas SQL sobre el motor relacional de la base de datos, el tiempo de respuesta puede aumentar considerablemente con el tamaño de la base de datos, lo que significa que, a mayor tamaño, más lenta será la respuesta.

Funcionalidad SQL restringida: Al utilizar principalmente sentencias SQL para interactuar con el motor relacional de la base de datos, ROLAP no puede satisfacer todas las necesidades de las consultas multidimensionales, ya que depende de las consultas para que soporte el motor de la base de datos. Sin embargo, recientemente se han implementado soluciones externas que permiten realizar formulaciones más complejas, lo que ayuda a mitigar algunas de estas limitaciones.

C. HOLAP (Hybrid online analytical processing).

Vitt y Luckevich (2002) describen HOLAP como una fusión de MOLAP y ROLAP. Aunque no se considera un método de almacenamiento de datos independiente, se refiere a la capacidad de distribuir datos entre bases de datos relacionales y multidimensionales, aprovechando las ventajas de ambos enfoques.

Tamayo y Moreno (2006) explican que HOLAP, que significa Hybrid Online Analytical Processing, emplea ambas técnicas: MOLAP ofrece análisis de datos agregados y métricas pre calculadas, mientras que ROLAP destaca por su escalabilidad, capacidad de generar informes en tiempo real, manejo simultáneo de múltiples usuarios y una gestión avanzada de datos.

Figura 11

Diferencias entre MOLAP, ROLAP, HOLAP

	MOLAP	ROLAP	HOLAP
Datos	Detalle y precalculados (agregados)	Detalle y agregados	Detalle y agregados
Estructura	Matrices comprimidas	Tablas relacionales	Multidimensional
Administración	Especialista en BDMD	Administrador BD	Administrador BD
Acceso	Lenguaje Especializado	SQL	SQL

Nota: Tomada de Tamayo M. y Moreno F., 2006.

2.5.8. Sistema gestor de base de datos (SGBD)

Nevado (n.d.) señala que un Sistema de Gestión de Bases de Datos (SGBD) cumple funciones esenciales relacionadas con la descripción, manipulación y control de los datos. Estas funciones incluyen: a) la definición de datos en distintos grados de abstracción (físico, lógico y externo), b) la manipulación de la información en el motor de la base de datos, que abarca la inserción, modificación, eliminación y consulta de datos, c) la garantía de la

consistencia de la base de datos, y d) la regulación de la privacidad y seguridad de los datos almacenados.

Gómez y De Abajo (1998) describen un SGBD como un conjunto de herramientas que facilitan a los usuarios la gestión de la información almacenada en una base de datos.

De acuerdo con Pons, Marín, Medina, Acid y Vila (2005), en un sistema gestor de bases de datos relacionales se pueden identificar dos tipos de tablas. Por un lado, están las tablas de usuario, que almacenan los datos operativos de la organización, y por otro, las tablas del sistema, que contienen información sobre la estructura de la base de datos, también conocida como el catálogo.

Osorio (2008) sostiene que los usuarios de un sistema de base de datos pueden realizar consultas sobre tablas, insertar nuevas tuplas o modificar y eliminar las existentes. Estas acciones se llevan a cabo mediante lenguajes conocidos como lenguajes de consulta relacional.

Cobo (n.d.) menciona que los lenguajes de álgebra relacional y cálculo relacional juegan un papel clave en el éxito de los Sistemas Gestores de Bases de Datos Relacionales (SGBDR) a nivel comercial.

Por su parte, Luque et al. (2002) indican que un SGBD está compuesto por un conjunto de aplicaciones que proporcionan a los usuarios las herramientas necesarias para realizar diversas actividades dentro de la base de datos, las actividades se detallan a continuación:

- a. Definir datos a distintos niveles de abstracción (físico, lógico y externo).
- b. Manipular la información en la base de datos, lo que implica inserciones, modificaciones, eliminaciones y consultas.
- c. Mantener la integridad de la base de datos, asegurando la validez de los datos y sus relaciones.
- d. Controlar la privacidad y seguridad de la información en la base de datos.

Estos medios son esenciales para establecer todas las características requeridas en una base de datos.

BASE DE DATOS

Según Oppel y Sheldon (2009), aunque no existe una definición única y universalmente aceptada, la mayoría de las definiciones coinciden en que una base de datos es, por lo menos una colección de datos organizados de manera estructurada, la cual se describe mediante metadatos que especifican dicha estructura. Se puede considerar a los metadatos como información que detalla cómo se almacenan los datos dentro de una base de datos.

Camps et al. (2005) afirman que una base de datos en un Sistema de Información representa de forma integrada los conjuntos de instancias de diferentes tipos de entidades del sistema, así como sus interrelaciones. Esta representación informática, que consiste en un conjunto estructurado de datos, debe ser accesible y utilizable de manera compartida por múltiples usuarios de diversas categorías.

BASE DE DATOS RELACIONAL

Según Osorio (2008), el modelo de una base de datos relacional, se ha establecido como un paradigma central en las tecnologías de la información, ganando aceptación y consolidación generalizadas. Este éxito ha facilitado la implementación efectiva del enfoque orientado a objetos en la actualidad.

Por otro lado, Heurtel (2009) describe las bases de datos relacionales como organizadas de acuerdo con el modelo propuesto por Edgar Codd en la década de 1970. Este modelo organiza los datos en tablas vinculadas lógicamente, donde cada tabla está formada por columnas que describen las filas que contiene. Las conexiones entre las tablas se realizan mediante columnas compartidas.

Según Oppel y Sheldon (2009), el componente esencial del modelo relacional es la relación, que se define como una estructura tabular de columnas y filas que representa una entidad específica compuesta por datos interrelacionados. Estas entidades pueden ser personas, lugares, objetos, eventos o conceptos sobre los cuales se recopilan datos, como autores, libros o transacciones de ventas. Cada relación contiene uno o más atributos, que son datos simples que describen o caracterizan la entidad de alguna manera.

2.5.9. Metodología Kimball

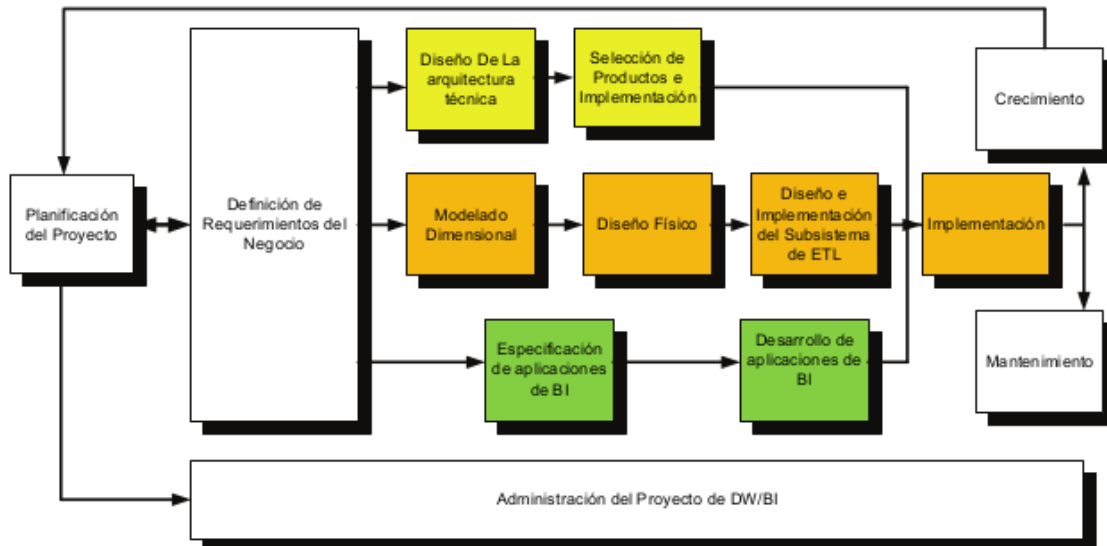
La metodología de Kimball incluía el modelado dimensional, que se convertirá en el estándar para el modelado de inteligencia de almacenamiento de datos empresariales. El modelado dimensional incluye tablas dinámicas con datos detallados y tablas con datos de búsqueda (Sherif, 2016, p. 10).

Los principios del Data Warehouse según Kimball, como se citó en Vela (2019) son:

- a. **Enfocarse en el negocio:** Es crucial identificar los requerimientos específicos del negocio y analizar cómo se relacionan sus diferentes áreas.
- b. **Construir una base de información adecuada:** La información generada debe ser completa y fácil de interpretar, reflejando con precisión las necesidades de la organización.
- c. **Entregas en intervalos con incrementos sustanciales:** Los Datamarts deben desarrollarse en plazos determinados, priorizando el orden de los incrementos, para ir construyendo gradualmente un Data Warehouse completo.
- d. **Ofrecer una respuesta integral:** Proveer a los usuarios los recursos necesarios, diseñando un Data Warehouse con datos de alta calidad, lo que permitirá la generación de informes mediante herramientas de inteligencia de negocios para un análisis adecuado en la toma de decisiones.

Figura 12

Ciclo de vida Kimball



Nota: Tomada de Rivadera, s.f.

A continuación, se describe el ciclo de desarrollo de un proyecto de Data Warehouse. En primer lugar, se debe planificar el proyecto, lo que implica establecer el propósito, los objetivos específicos y el alcance. En segundo lugar, se identifican los requerimientos del negocio, basados en entrevistas y análisis de documentos internos, para comprender los procesos organizacionales. Esto incluye las necesidades de información, datos y la creación de una matriz bus que relacione procesos y objetos empresariales. Posteriormente, se lleva a cabo el modelado dimensional, seleccionando los procesos de negocio más relevantes, estableciendo el nivel de granularidad de los indicadores, definiendo dimensiones y atributos, y asignando las medidas correspondientes en las tablas de hechos.

A continuación, se diseña el entorno físico de la base de datos, considerando los tipos de datos y aspectos de seguridad. Luego, se diseña e implementa el subsistema ETL, que abarca la extracción, transformación y carga de los datos en el modelo relacional. En paralelo, se diseña la arquitectura técnica, teniendo en cuenta los requerimientos empresariales, el entorno tecnológico y la estrategia organizacional.

La selección de productos e implementación es el siguiente paso, en el cual se eligen los componentes de hardware, la base de datos y las herramientas necesarias, que posteriormente se instalan y prueban. Seguido a esto, se especifican las aplicaciones de inteligencia de negocios (BI), asignando roles o perfiles para los distintos niveles de análisis de los usuarios. Posteriormente, se desarrolla la aplicación de BI, enfocada en la creación de reportes específicos, y se lleva a cabo la implementación, proporcionando acceso a los usuarios finales tras haber recibido capacitaciones adecuadas.

El proyecto se administra mediante un monitoreo constante para asegurar que se cumplan los requerimientos de la organización. Finalmente, se enfoca en el mantenimiento y crecimiento del sistema, lo que implica la incorporación de innovaciones y mejoras en la experiencia del usuario final.

Capítulo III

Material y métodos

3.1. Tipo de investigación

Para Supo (2015), los estudios observacionales, en los que el investigador no interviene para alterar los resultados de las mediciones, incluso cuando él mismo las realiza, permiten que los datos obtenidos y la información registrada reflejen el estado natural de las unidades de estudio.

Para Supo (2015), los estudios retrospectivos se llevan a cabo utilizando datos previamente registrados. Emplean información que proviene de registros no controlados, en los cuales el investigador no tuvo participación.

Para Supo (2015), los estudios transversales son aquellos en los que todas las variables se miden en una única ocasión, sin importar el tiempo que se necesite para completar las mediciones de todos los elementos del grupo de estudio.

3.2. Nivel de investigación

Según Bernal (2010), la investigación descriptiva se caracteriza por su capacidad para identificar y detallar las principales características del objeto de estudio. Este tipo de investigación suele apoyarse en diversas técnicas, como encuestas, entrevistas, observación y revisión de documentos. Entre los temas comunes que abarca se encuentran los estudios diagnósticos, la creación de guías, modelos, prototipos, estudios de mercado, la descripción de comportamientos, y los análisis de tiempo y movimiento.

Por su parte, Supo (2015) señala que la investigación descriptiva representa el primer nivel dentro de los estudios cuantitativos, donde el uso del análisis estadístico es esencial para realizar cálculos que ayuden a alcanzar los objetivos del estudio. Este tipo de estudios suele ser "univariado", ya que el análisis se enfoca en una sola variable, lo que justifica que el nivel de investigación sea descriptivo.

3.3. Diseño de la investigación

De acuerdo con el autor Hernández Sampieri et. al. (2014), se puede definir la investigación no experimental, "como estudios realizados sin manipulación intencional de variables, estos

son estudios en los que no cambiamos intencionalmente las variables independientes para ver su efecto en otras variables. En la investigación no experimental, observamos los fenómenos tal como ocurren en su contexto natural y los analizamos después”.

3.4. Población y muestra

3.4.1. Población

La población está conformada por todos los registros académicos en la Universidad Nacional de San Cristóbal de Huamanga, durante el 2024.

3.4.2. Muestra

No hay muestra, ya que se estudiará toda la población, así que se hará un censo.

3.5. Variables e indicadores

3.5.1. Definición conceptual de las variables

Primera variable de interés

Datamart: Un Datamart es un método sencillo de almacenamiento de datos, enfocado en un tema o área de negocio específica para facilitar el análisis dentro de la organización.

Variables descriptivas de la primera variable de interés

Diseño y arquitectura: El diseño y la arquitectura del Datamart son esenciales para garantizar su funcionalidad y eficiencia. Un Datamart está diseñado para ser un repositorio especializado que puede realizar consultas rápidas y análisis de datos detallados para áreas específicas.

Integración de datos: La integración de datos es el proceso de combinar información de varias fuentes. Es fundamental para el correcto funcionamiento del Datamart, ya que permite la unificación de datos de sistemas dispares en un solo repositorio centralizado, mejorando la calidad de los datos y facilitando su análisis.

Calidad de los datos: La calidad de los datos en un Datamart es crucial para asegurar la precisión y la utilidad de los informes generados.

Segunda variable de interés

Seguimiento académico: Es el proceso académico que tiene como resultado analizar el rendimiento académico de los estudiantes, conllevan a tomar decisiones al respecto con énfasis en la calidad de la educación.

VARIABLES DESCRIPTIVAS DE LA SEGUNDA VARIABLE DE INTERÉS

Eficiencia del seguimiento: La eficiencia se mide por la capacidad del sistema para procesar y analizar datos rápidamente, permitiendo la toma de decisiones en tiempo real. Un Datamart puede reducir significativamente el tiempo necesario para generar reportes académicos.

Precisión en la Información: La precisión de la información es clave para la efectividad del seguimiento académico. Un Datamart con datos precisos permite generar reportes que reflejan de manera exacta el rendimiento de los estudiantes, evitando decisiones basadas en datos incorrectos.

Accesibilidad de los Datos: La accesibilidad es crucial para que los usuarios del sistema puedan consultar y obtener información académica de manera fácil y eficiente.

3.5.2. Definición operacional de las variables

Primera variable de interés

X: Datamart

VARIABLES DESCRIPTIVAS

X1: Diseño y arquitectura

X2: Integración de datos

X3: Calidad de los datos

Segunda variable de interés

Y: Seguimiento académico

VARIABLES DESCRIPTIVAS

Y1: Eficiencia del seguimiento.

Y2: Precisión en la Información

Y3: Accesibilidad de los Datos

3.6. Técnicas e instrumentos para el tratamiento de datos e información

3.6.1. Técnicas para recolectar información

Para obtener los datos necesarios, se emplearon dos técnicas principales: entrevistas y análisis de documentos.

3.6.2. Instrumentos para recolectar información

Se utilizó la guía de entrevistas para recopilar la información proporcionada por la persona encargada del procesamiento de datos en la oficina de tecnologías de la información. Además, se utilizó una ficha de análisis de la base de datos, la cual permitió extraer información relevante de la base de datos de la Universidad Nacional de San Cristóbal de Huamanga.

3.6.3. Herramientas para el tratamiento de datos e información

Las tecnologías usadas fueron:

Tabla 1

Instrumentos tipos de aplicaciones utilizadas para el manejo de los datos

APLICACIÓN	PROVEEDOR	SOPORTE
Windows 11	Corporación Microsoft	La última iteración de Windows, producto de sistemas operativos producida por la empresa Microsoft Corporation.
Microsoft Sql Server 2019	Corporación Microsoft	Esta aplicación proporciona una base de datos robusta para la instalación tanto del servidor web como del servidor de datos.
Visual Studio 2022	Corporación Microsoft	Es un Aplicación para desarrollar múltiples proyectos. Soporta varios lenguajes de programación.
Power BI	Corporación Microsoft	es un servicio de análisis de datos de Microsoft que ofrece visualizaciones interactivas y capacidades de inteligencia empresarial (también conocida

como BI). La interfaz es lo suficientemente simple como para que los usuarios puedan crear informes y paneles por sí mismos.

Nota: Tabla de elaboración propia.

3.7. Técnicas para aplicar la metodología Kimball

Como se detalla en la revisión de la literatura desarrollada en el Capítulo II, Sección 2.5.9, se desarrolla el proceso que contiene las fases para el desarrollo del Datamart, aplicando la metodología Kimball. Estas fases se presentan en las siguientes tablas.

Tabla 2

Programación del proyecto del Datamart

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Organizar las actividades necesarias para la creación del Datamart	Planificación de las tareas	Disponibilidad del recurso. Dificultad del plan	Especialista

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 3

Determinar los requerimientos

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Especificar los requisitos funcionales y no funcionales	Síntesis de los requisitos funcionales y no funcionales.	Consulta con expertos del negocio. Reunión con los usuarios. Revisión documental de los procesos.	Especialista del negocio.

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 4*Diseño de la arquitectura técnica*

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Diseñar la arquitectura técnica	Arquitectura técnica	Revisión de los requisitos funcionales. Revisión del entorno tecnológico.	Especialista del negocio

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 5*Selección de producto*

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Selección de producto	Herramientas	Software Hardware	Especialista Desarrollador

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 6*Modelo dimensional*

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Realizar el modelo dimensional	Modelo dimensional	Analizar el proceso de negocio escogido Elección del objetivo de las tablas de hechos	Diseñador Especialista
Diseñar las tablas de dimensiones	Esquema de las tablas de dimensiones	Determinar el nombre de la tabla dimensión Agregar un campo de la clave principal Replantear nombres de campos	Especialista del negocio, desarrollador
Realizar las tablas de hechos	Esquema de las tablas de hechos	Determinar el nombre de la tabla de hechos Definir la clave primaria	Especialista del negocio, desarrollador

		Agregar campos de la tabla de hechos como indicadores	
Hacer uniones	Esquema del esquema de uniones	Esquema estrella	Desarrollador
Realizar el modelo lógico	Esquema del modelo lógico	Esquema estrella	Desarrollador

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 7

Diseño físico

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Desarrollar el modelo físico	Diseño físico	Transformar el modelo lógico a modelo físico. Formular la consulta SQL desde el modelo físico.	Diseñador Analista

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 8

Proceso ETL

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Extracción, transformación de la carga inicial	Carga del almacén intermedio	Limpieza de datos Procesos ETL	Desarrollador
Actualización de las tablas	Datos cargados en el Datamart	Carga de todas las tablas y dimensiones	Analista de negocio, desarrollador
Creación de cubos multidimensionales	Cubo multidimensional	Creación de indicadores, atributos y jerarquías.	Desarrollador

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 9*Implementación*

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Implementación	Desarrollar el cubo multidimensional	Generación de indicadores, atributos y jerarquías.	Desarrollador

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 10*Especificación de aplicaciones para usuarios finales*

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Especificación de aplicación para usuarios finales	Roles	Identificación de la parte operativa y estratégica	Especialista

Nota: Elaborado a partir de Robles y Girón, 2015.

Tabla 11*Mantenimiento y crecimiento*

ACTIVIDAD	ENTREGABLE	PROCEDIMIENTO	COORDINADOR
Mantenimiento y crecimiento	Mantenimiento y crecimiento	Identificación de nuevos requerimientos	Analista Desarrollador Especialista del negocio

Nota: Elaborado a partir de Robles y Girón, 2015.

Capítulo IV

Resultados de la investigación

4.1. Resultados aplicando la metodología Kimball

4.1.1. Planificación del proyecto

Acorde a la metodología Kimball, detallados en las tablas 2 al 11, donde se muestran las fases para el desarrollo de la metodología, descritos en el capítulo II sección 2.5.9, nos indica realizar la planificación del proyecto para poder tener un mejor control de las actividades a desarrollarse.

4.1.2. Definición de los requerimientos

Para la definición de los requerimientos se analizó los documentos disponibles de la oficina de tecnologías de la información, con el objetivo de elaborar el Datamart de seguimiento académico que permita centralizar las fuentes de información y optimizar el proceso de seguimiento académico de la universidad nacional de san Cristóbal de huamanga.

4.1.2.1. Requerimientos funcionales

Tabla 12

Lista de requerimientos

Nro.	INTERROGANTE	OBJETIVO
1	¿Cuántos matriculados hay en un periodo por género, escuela y facultad?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.
2	¿Cantidad de veces llevo un alumno sus cursos?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.
3	¿Cantidad de aprobados por curso y escuela durante un periodo?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.

4	¿Cantidad de desaprobados por curso y escuela durante un periodo?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.
5	¿Cantidad de aprobados por facultad durante un periodo?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.
6	¿Cantidad de desaprobados facultad durante un periodo?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.
7	¿Cuáles son las notas por curso y periodos académicos cursados por el estudiante?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.
8	¿Cuáles es la tasa de éxito de un curso por periodo?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.
9	¿Cuáles es la tasa de fracaso de un curso por periodo?	Conocer la cantidad permitirá realizar un monitoreo eficiente de los indicadores y emitir reportes oportunos.

Nota: Tabla de elaboración propia.

4.1.2.2. Requerimientos no funcionales

- a. El Datamart debe actualizar automáticamente los datos de rendimiento académico extraídos de la base de datos principal de la universidad, sincronizando cualquier cambio o actualización realizada en la fuente original.
- b. El Datamart debe ser escalable, permitiendo agregar más datos académicos (nuevas asignaturas, periodos académicos, cursos) sin comprometer el rendimiento.

- c. El Datamart debe ofrecer tiempos de respuesta rápidos en la consulta de datos y generación de reportes, asegurando que los usuarios puedan acceder a la información en tiempo real.
- d. El software utilizado para mostrar los reportes finales a los usuarios es el Microsoft Power BI.

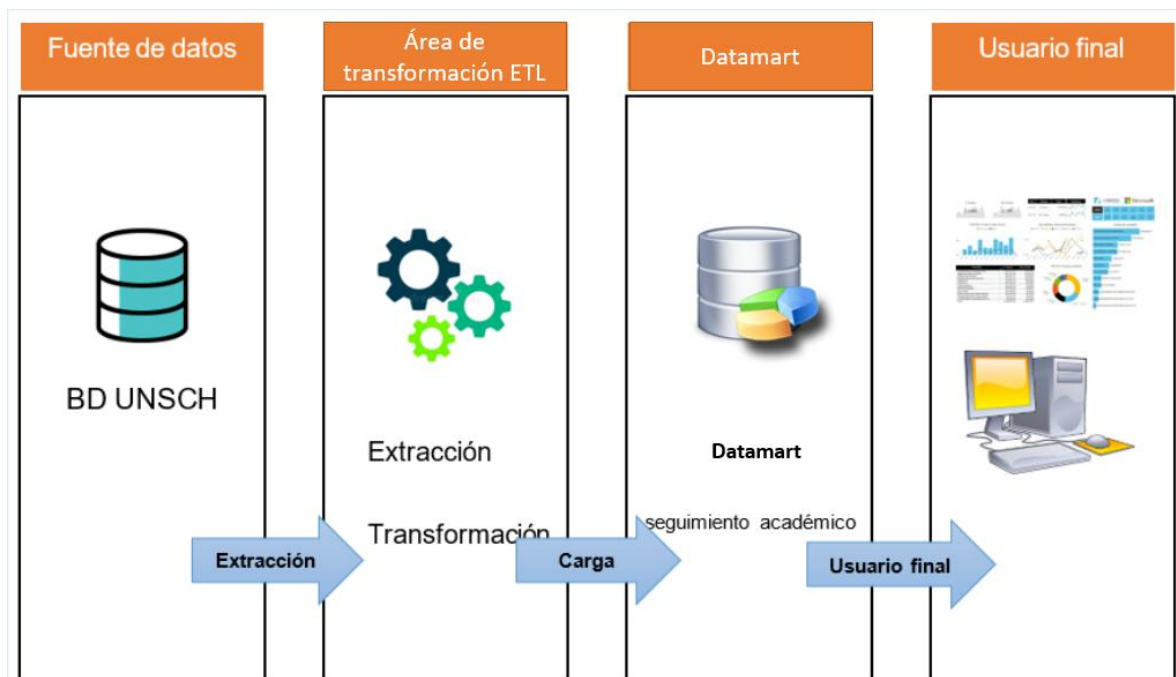
4.1.3. Diseño de la arquitectura técnica

Nivel de datos

En esta parte se presenta la arquitectura técnica necesaria para el desarrollo del Datamart para la Universidad Nacional De San Cristóbal De Huamanga.

Figura 13

Arquitectura técnica



Composición de la arquitectura técnica en los siguientes niveles:

- a. Fuente de datos:** Fuente de datos principal de la universidad para obtener los datos necesarios para la implementación del Datamart de seguimiento académico.
- b. Área de transformación ETL:** En esta parte se hace la transformación ETL, donde se extraen los datos de las fuentes de datos, estos se transforman para finalmente hacer la carga en el Datamart de seguimiento académico.

c. Datamart: El Datamart de seguimiento académico nos permitirá sacar reportes y consultas necesarios para cumplir con la implementación del Datamart según lo requerido.

d. Usuario final: Son las personas interesadas e involucradas en la parte final del trabajo que podrán consumir los datos organizados y que podrán ser mostrados a través de Microsoft Power BI.

4.1.4. Selección del producto

Hardware

Para los servidores y la base de datos se necesitará las siguientes características:

Disco duro: 1 Tb.

Memoria Ram: 12 Gb.

Procesador Intel Core I5

Software

Tabla 13

Software para la construcción del Datamart

PROCESOS	HERRAMIENTA
Procesos ETL	SQL server Busines Intelligence
Cubo	Analysis Service
Datamart	SQL Server 2019
Interfaz	Microsoft Power BI

Nota: Tabla de elaboración propia.

4.1.5. Modelo dimensional

4.1.5.1. Desarrollo del modelo dimensional

A continuación, se detalla el desarrollo del modelo dimensional en cuatro etapas.

a. Elección del Datamart

En esta etapa se va a elegir el Datamart de seguimiento académico para la Universidad Nacional De San Cristóbal De Huamanga, el cual será utilizado para generar consultas y reportes que apoyen en el seguimiento académico de los estudiantes. Esto permitirá a las autoridades tomar decisiones estratégicas que incrementen las tasas de éxito en la aprobación de cursos y favorezcan la culminación adecuada de la malla curricular en las diferentes escuelas.

Se usará la base de datos proporcionada por la universidad para extraer los datos y procesarlos, eligiendo las dimensiones y hechos que nos permitan desarrollar el Datamart para lo cual nos apoyaremos en los requisitos definidos en la tabla 12 del capítulo IV.

Tabla 14

Lista de indicadores

Nro.	INDICADORES
IND 1	Cantidad de matriculados por género, escuela y facultad.
IND 2	Cantidad de veces que llevo un alumno sus cursos.
IND 3	Cantidad de aprobados por curso y escuela durante un periodo.
IND 4	Cantidad de desaprobados por curso y escuela durante un periodo.
IND 5	Cuál es el consolidado de alumnos por curso y escuela.
IND 6	Cantidad de aprobados por facultad durante un periodo.
IND 7	Cantidad de desaprobados por facultad durante un periodo.
IND 8	Cuáles son las notas por curso y ciclos académicos cursados por el estudiante.
IND 9	Cuáles es la tasa de éxito de un curso por periodo.
IND 10	Cuáles es la tasa de fracaso de un curso por periodo.

Nota: Tabla de elaboración propia.

Tabla 15

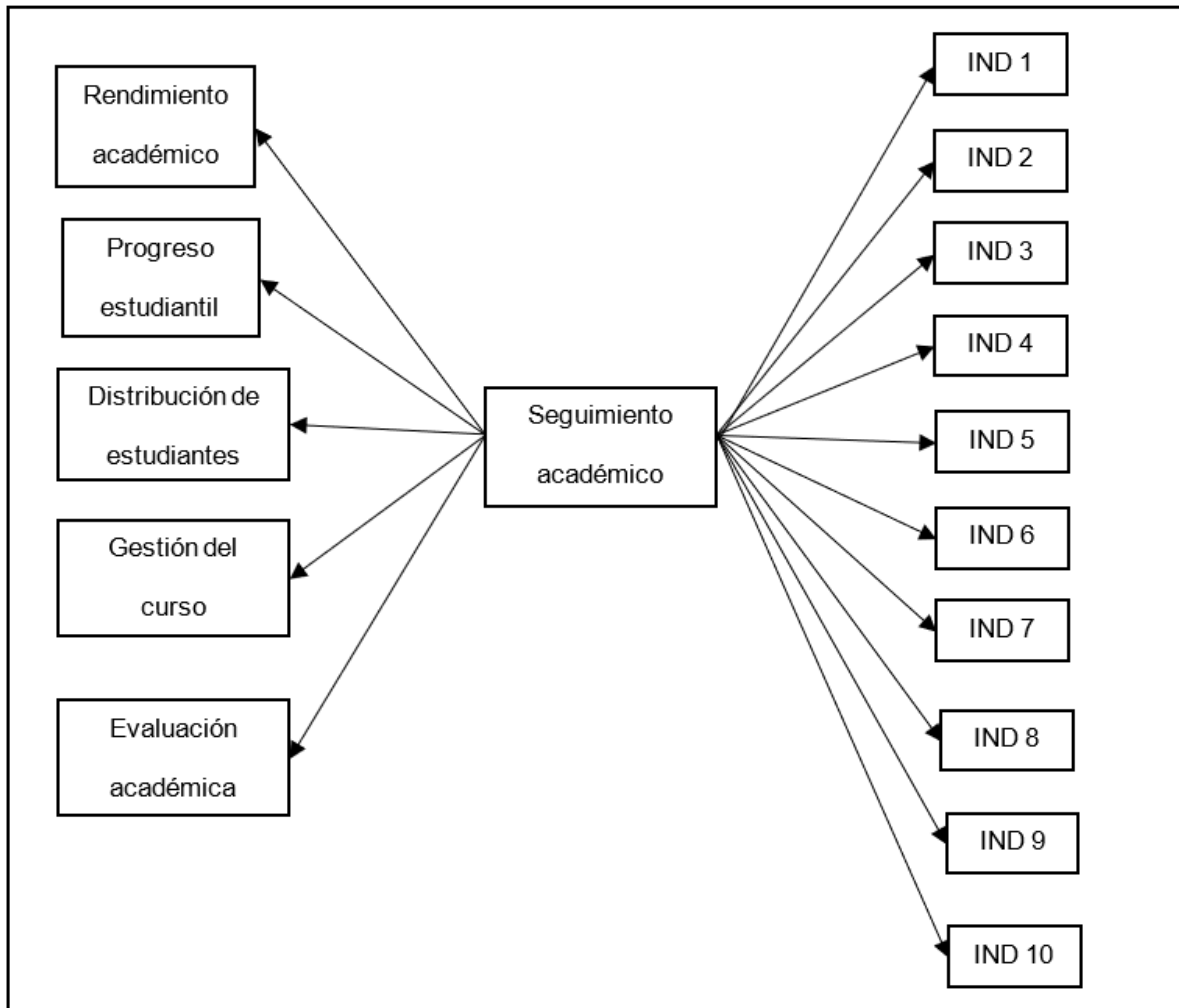
Lista de perspectivas

Nro.	Perspectivas
1	Rendimiento académico.
2	Progreso estudiantil.
3	Distribución de estudiantes.
4	Gestión del curso.
5	Evaluación académica.

Nota: Tabla de elaboración propia.

Figura 14

Diagrama del modelo conceptual inicial



b. Elección del objetivo de la tabla de hechos

En esta fase se define el objetivo de la tabla de hechos, se va definir los registros que contemplen la tabla de hechos del Datamart de seguimiento académico.

Tabla 16

Objetivo de la tabla de hechos

TABLA DE HECHO	OBJETIVO
Seguimiento académico	Contiene las medidas que se analizarán, como calificaciones, aprobaciones, desaprobaciones, repeticiones de cursos, entre otros.

Nota: Tabla de elaboración propia.

c. Elección de las dimensiones

Para la elección de las dimensiones Kimball, nos menciona que se debe elegir adecuadamente las asociaciones con la tabla de hechos, se muestra en la siguiente tabla.

Tabla 17

Dimensiones de la tabla de hechos

TABLA DE HECHO	OBJETIVO	DIMENSIONES
Seguimiento académico	Contiene las medidas que se analizarán, como calificaciones, aprobaciones, desaprobaciones, repeticiones de cursos, entre otros.	<ul style="list-style-type: none">- Estudiante- Curso- Facultad/Escuela- Periodo académico

Nota: Tabla de elaboración propia.

Para seleccionar las dimensiones, se ha considerado la necesidad específica del Datamart, se realizó el modelado, como se define previamente en las tablas 14 y 15.

d. elección de la tabla hecho (Fact)

Cada tabla de hechos tiene como objetivo definir claramente los datos que se desean medir y establecer el alcance que deben cubrir. Se ah creado una sola tabla de hechos para el seguimiento académico en este caso. Esta tabla facilitará la toma de decisiones al capturar y analizar datos importantes sobre el rendimiento académico de los estudiantes.

Tabla 18

Elección de hechos

TABLA DE HECHO	HECHOS
Seguimiento académico	- Calificación - Estado del curso - Veces cursado

Nota: Tabla de elaboración propia.

4.1.5.2. Diagramas de las tablas de dimensiones

En este apartado se desarrollará de manera detallada las dimensiones ya elegidas en la tabla 17.

a. Dimensión Facultad/Escuela

Figura 15

Jerarquía de atributos de la dimensión EscuelaFacultad

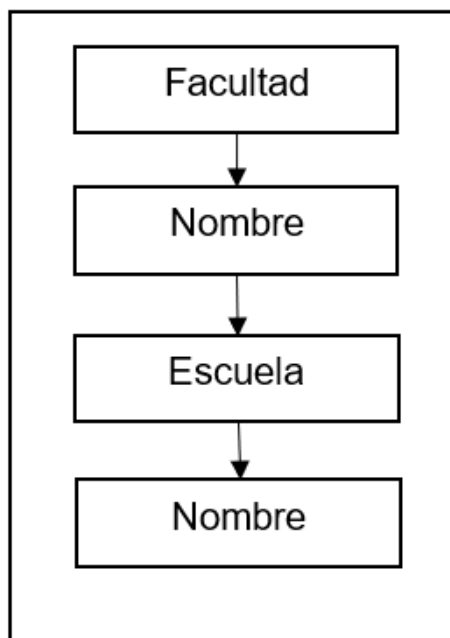


Tabla 19

Dimensión Facultad/Escuela

NOMBRE DEL ATRIBUTO	DESCRIPCIÓN DEL ATRIBUTO
Fe_IdFacultad	Representa el identificador de la facultad
Fe_NombreFacultad	Representa el nombre de la facultad
Fe_IdEscuela	Representa el identificador de la facultad
Fe_NombreEscuela	Representa el nombre de la facultad

Nota: Tabla de elaboración propia.

b. Dimensión Estudiante

Figura 16

Jerarquía de atributos de la dimensión Estudiante

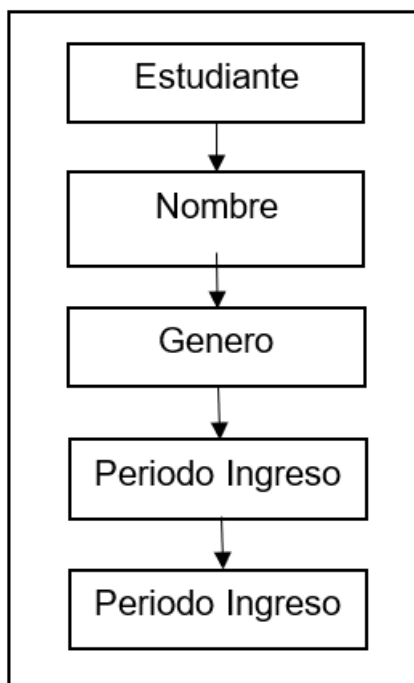


Tabla 20

Dimensión Estudiante

NOMBRE DEL ATRIBUTO	DESCRIPCIÓN DEL ATRIBUTO
Es_IdEstudiante	Representa el identificador del estudiante
Es_NombreEstudiante	Representa el nombre del estudiante
Es_Genero	Representa el género del estudiante
Es_PeriodoIngreso	Representa el periodo de ingreso del estudiante

Nota: Tabla de elaboración propia.

c. Dimensión Curso

Figura 17

Jerarquía de atributos de la dimensión Curso

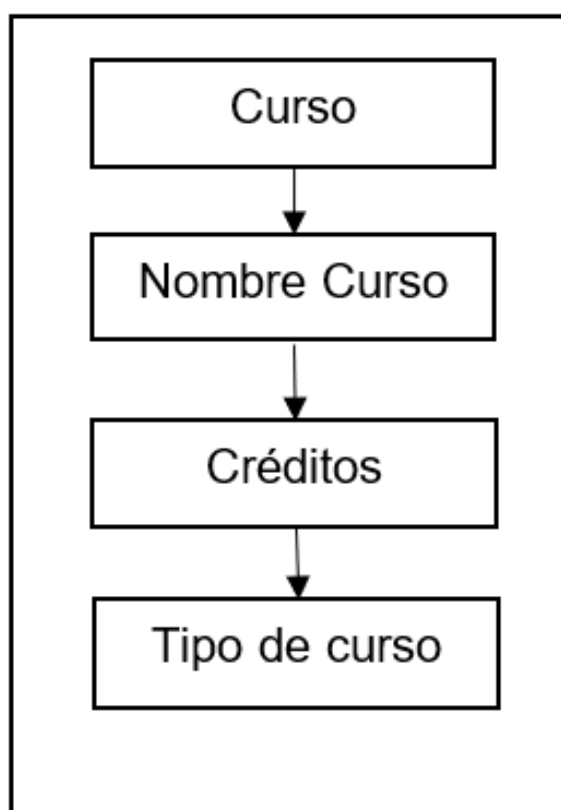


Tabla 21

Dimensión Curso

NOMBRE DEL ATRIBUTO	DESCRIPCIÓN DEL ATRIBUTO
Cu_IdCurso	Representa el identificador del curso
Cu_NombreCurso	Representa el nombre del curso
Cu_Creditos	Representa los créditos del curso
Cu_TipoCurso	Representa el tipo del curso

Nota: Tabla de elaboración propia.

d. Dimensión Periodo Académico

Figura 18

Jerarquía de atributos de la dimensión Periodo Académico

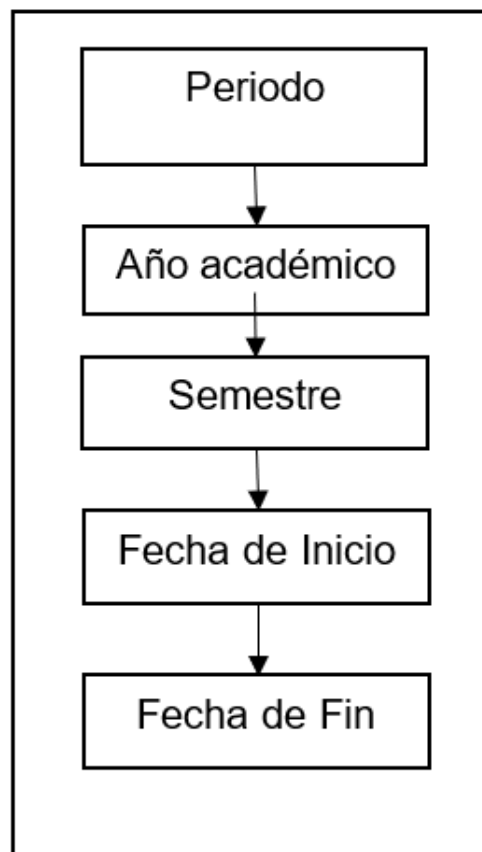


Tabla 22

Dimensión Periodo académico

NOMBRE DEL ATRIBUTO	DESCRIPCIÓN DEL ATRIBUTO
Pa_IdPeriodoAcademico	Representa el identificador del periodo académico
Pa_AnioAcademico	Representa el año académico
Pa_Semestre	Representa el semestre del periodo académico
Pa_FechaInicio	Representa la fecha de inicio
Pa_FechaFin	Representa la fecha final

Nota: Tabla de elaboración propia.

e. Dimensión Plan Estudio

Figura 19

Jerarquía de atributos de la dimensión Plan Estudio

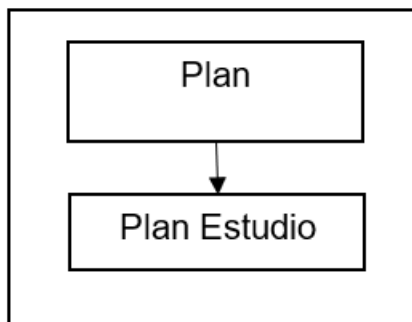


Tabla 23

Dimensión Plan Estudio

NOMBRE DEL ATRIBUTO	DESCRIPCIÓN DEL ATRIBUTO
P_IdPlan	Representa el identificador del plan de estudio
P_PlanEstudio	Representa el nombre del plan de estudio

Nota: Tabla de elaboración propia.

f. Dimensión Matricula

Figura 20

Jerarquía de atributos de la dimensión Matricula

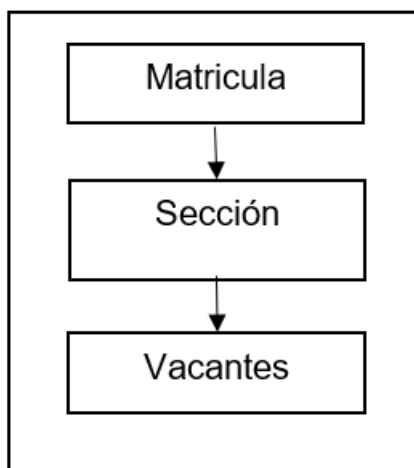


Tabla 24

Dimensión Matricula

NOMBRE DEL ATRIBUTO	DESCRIPCIÓN DEL ATRIBUTO
M_IdMatricula	Representa el identificador de la matricula
M_seccionMatricula	Representa la sección de la matricula
M_Vacantes	Representa las vacantes de la matricula

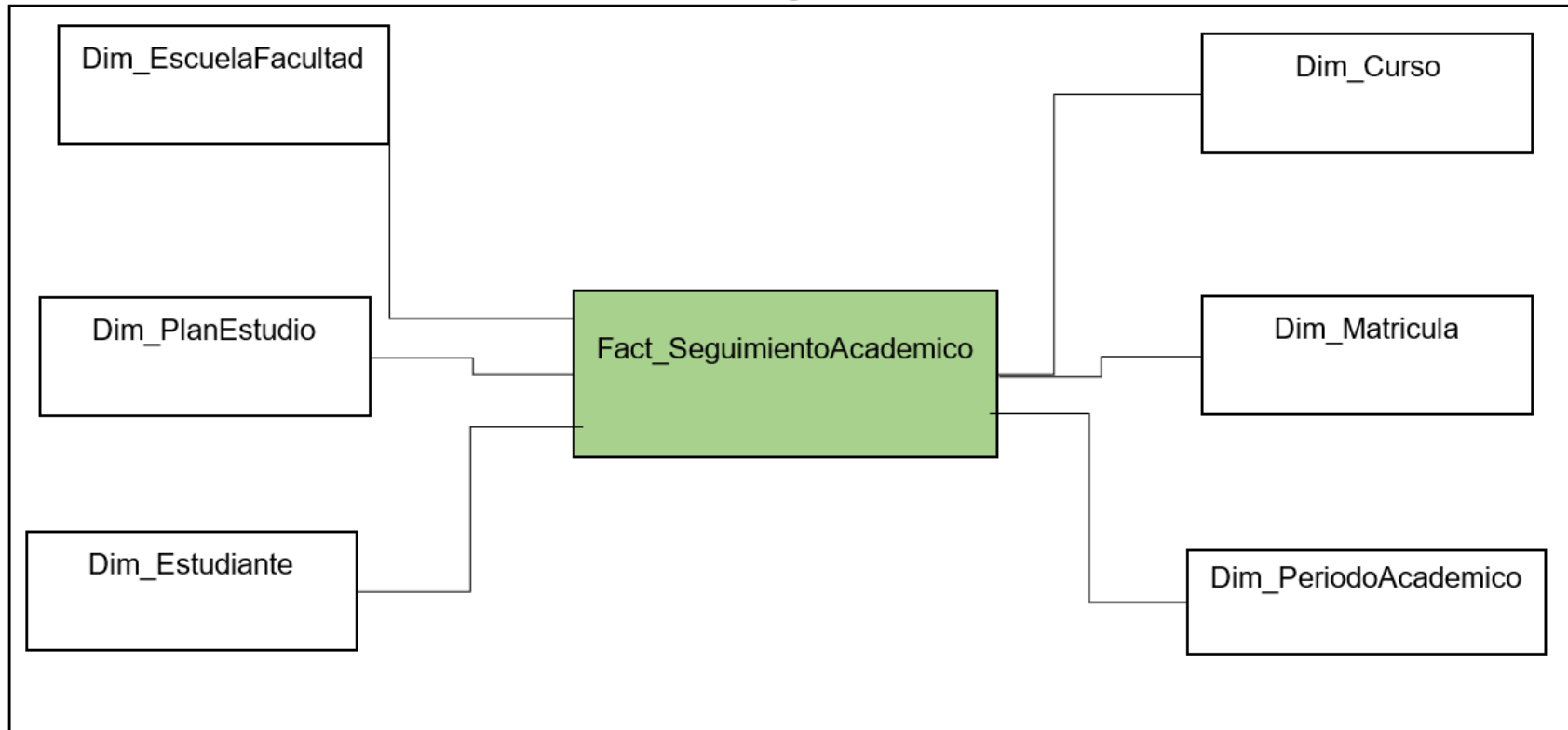
Nota: Tabla de elaboración propia.

4.1.5.3. Diagrama de la tabla de hecho

A continuación, se muestra el diagrama el diagrama lógico para la tabla de hecho con sus dimensiones, según la metodología desarrollada.

Figura 21

Diagrama de la tabla de hechos



Detalle de las dimensiones

Tabla 25

Detalle de las llaves de las dimensiones

NOMBRE DE LA TABLA	NOMBRE DE LA COLUMNA	DESCRIPCIÓN DE LA COLUMNA
Dimensión Facultad Escuela	Fe_IdFacultad	Llave primaria para la dimensión facultad/escuela, llave foránea para la tabla de hecho.
Dimensión Estudiante	Es_IdEstudiante	Llave primaria para la dimensión estudiante, llave foránea para la tabla de hecho.
Dimensión Curso	Cu_IdCurso	Llave primaria para la dimensión curso, llave foránea para la tabla de hecho.
Dimensión Periodo académico	Pa_IdPeriodoAcademico	Llave primaria para la dimensión periodo académico, llave foránea para la tabla de hecho.
Dimensión Plan	P_IdPlan	Llave primaria para la dimensión plan, llave foránea para la tabla de hecho.
Dimensión Matricula	M_IdMatricula	Llave primaria para la dimensión matricula, llave foránea para la tabla de hecho.

Nota: Tabla de elaboración propia.

Tabla 26

Detalles de las medidas de la tabla de hecho Fact_SeguimientoAcademico

NOMBRE DE LA TABLA	NOMBRE DE LA COLUMNA	DESCRIPCIÓN DE LA COLUMNA
Fact seguimiento académico	Calificación	Nota final del estudiante.
	Estado del curso	Indica si el estudiante aprobó, desaprobó o se retiró del curso.
	Veces cursado	Número de veces que el estudiante llevo el curso.

Nota: Tabla de elaboración propia.

Figura 22

Tabla de hecho del Datamart

Fact SeguimientoAcademico
Fe_IdEscuela
Es_IdEstudiante
Cu_IdCurso
Pa_IdPeriodoAcademico
P_IdPlan
M_IdMatricula
Calificacion
EstadoCurso
VecesCursado

4.1.5.4. Diagrama del modelo lógico

Una vez definidas las tablas de dimensiones y la tabla de hechos, y establecidas las relaciones entre ellas mediante las claves primarias, se han orientado estas estructuras hacia el diseño del Datamart. Para facilitar su visualización, se presenta el modelo lógico en un esquema estrella.

A continuación, se proporciona un listado detallado de todos los componentes que intervienen en el esquema estrella, lo cual permitirá una comprensión más clara y precisa de la estructura y funcionalidad del Datamart.

Modelado del esquema estrella

Tabla de dimensión Facultad/Escuela
Tabla de dimensión Estudiante
Tabla de dimensión Curso
Tabla de dimensión Periodo académico

Llaves primarias

Fe_IdFacultad

Es_IdEstudiante

Cu_IdCurso

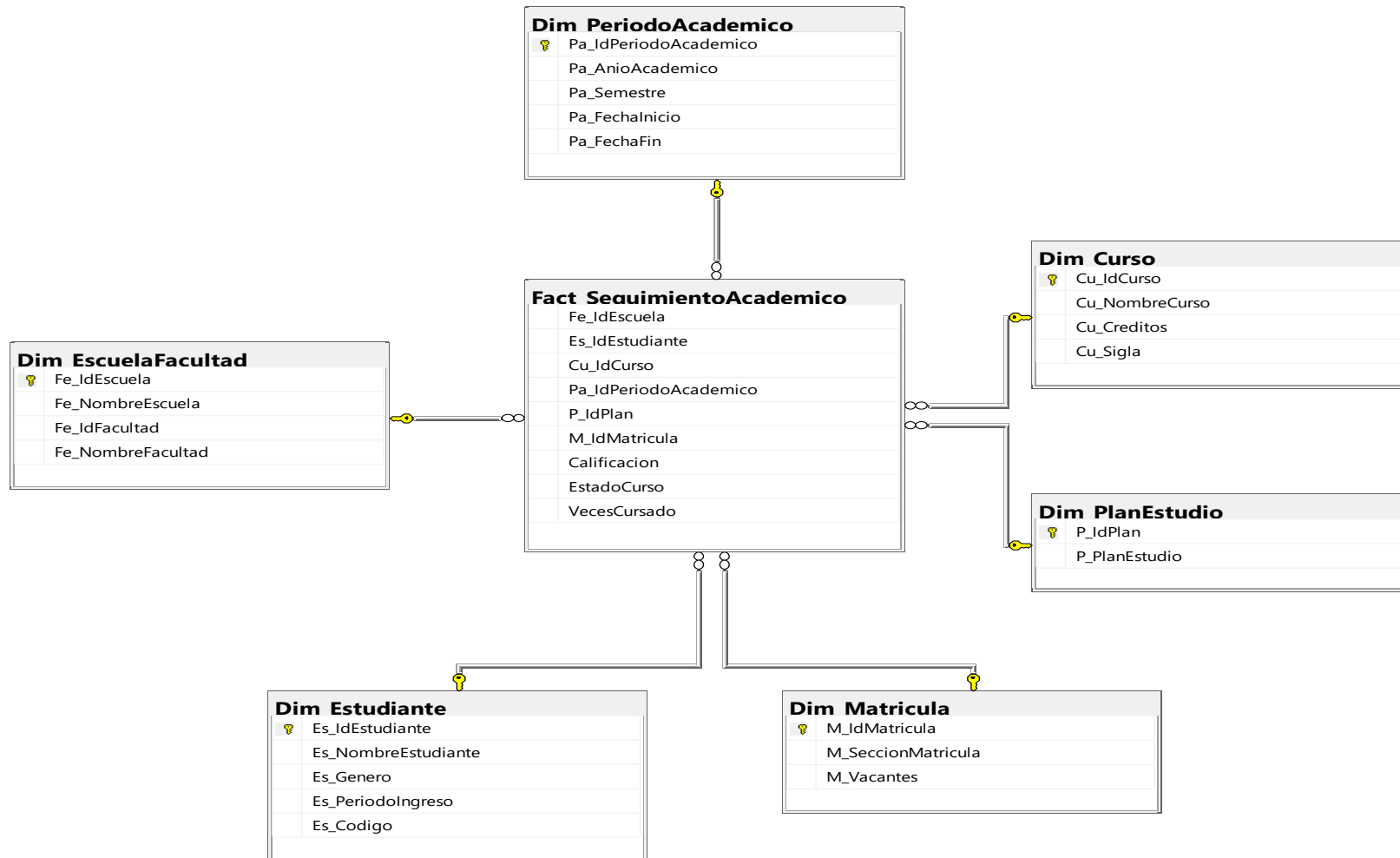
Pa_IdPeriodoAcademico

P_IdPlan

M_IdMatricula

Figura 23

Modelo lógico de Datamart



4.1.6. Diseño del modelo físico

En esta etapa ya habiendo desarrollado el modelo lógico en el esquema estrella, se asocian las dimensiones a la tabla de hechos, en el modelo físico se agregarán el tamaño de las variables de cada dimensión para ejecutar en el motor de la base de datos.

i. Tabla: Dim_Estudiante

Tabla 27

Tipo de datos para Dim_Estudiante

NOMBRE DEL CAMPO	TIPO DE VARIABLE	LONGITUD
Es_IdEstudiante	Char	36
Es_NombreEstudiante	Varchar	120
Es_Genero	Char	1
Es_PeriodoIngreso	Varchar	20
Es_Codigo	Varchar	50

Nota: Tabla de elaboración propia.

ii. Tabla: Dim_FacultadEscuela

Tabla 28

Tipo de datos para Dim_EscuelaFacultad

NOMBRE DEL CAMPO	TIPO DE VARIABLE	LONGITUD
Fe_IdEscuel	Char	36
Fe_NombreEscuela	Varchar	70
Fe_IdFacultad	Char	6
Fe_NombreFacultad	Varchar	70

Nota: Tabla de elaboración propia.

iii. Tabla: Dim_Curso

Tabla 29

Tipo de datos para Dim_Curso

NOMBRE DEL CAMPO	TIPO DE VARIABLE	LONGITUD
Cu_IdCurso	Char	36
Cu_NombreCurso	Varchar	120
Cu_Creditos	Decimal	(18,2)
Cu_Sigla	Varchar	50

Nota: Tabla de elaboración propia.

iv. Tabla: Dim_PeriodoAcademico

Tabla 30

Tipo de datos para Dim_PeriodoAcademico

NOMBRE DEL CAMPO	TIPO DE VARIABLE	LONGITUD
Pa_IdPeriodoAcademico	Char	36
Pa_AnioAcademico	Varchar	50
Pa_semestre	Varchar	50
Pa_Fechnicio	Datetime	
Pa_FechaFin	Datetime	

Nota: Tabla de elaboración propia.

v. Tabla: Dim_PlanEstudio

Tabla 31

Tipo de datos para Dim_PlanEstudio

NOMBRE DEL CAMPO	TIPO DE VARIABLE	LONGITUD
P_IdPlan	Char	36
P_PlanEstudio	Varchar	50

Nota: Tabla de elaboración propia.

vi. Tabla: Dim_Matricula

Tabla 32

Tipo de datos para Dim_Matricula

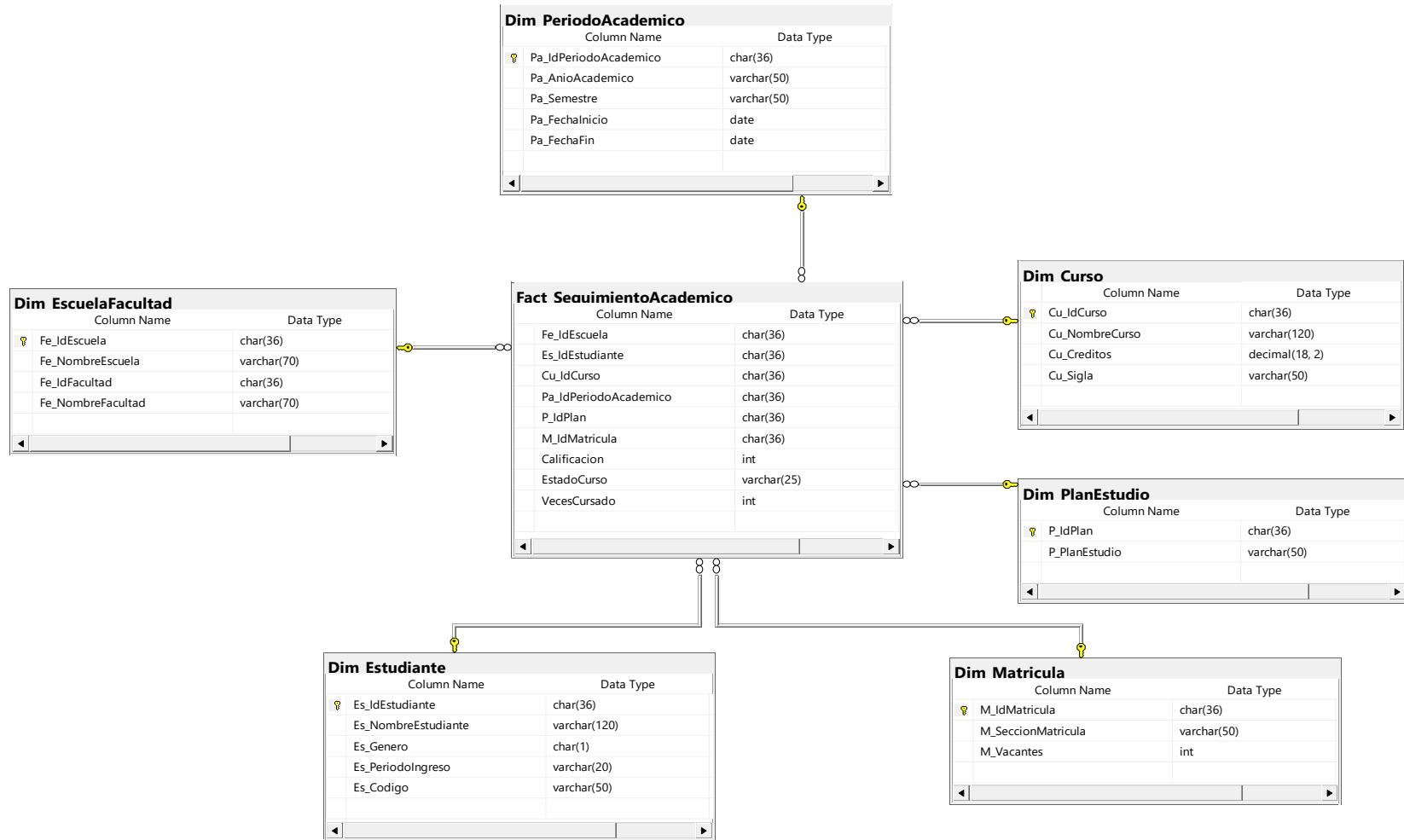
NOMBRE DEL CAMPO	TIPO DE VARIABLE	LONGITUD
M_IdMatricula	Char	36
M_SeccionMatricula	Varchar	50
M_Vacantes	Int	

Nota: Tabla de elaboración propia.

A continuación, se muestra el siguiente modelo físico para la base de datos del Datamart.

Figura 24

Diseño del modelo físico del Datamart

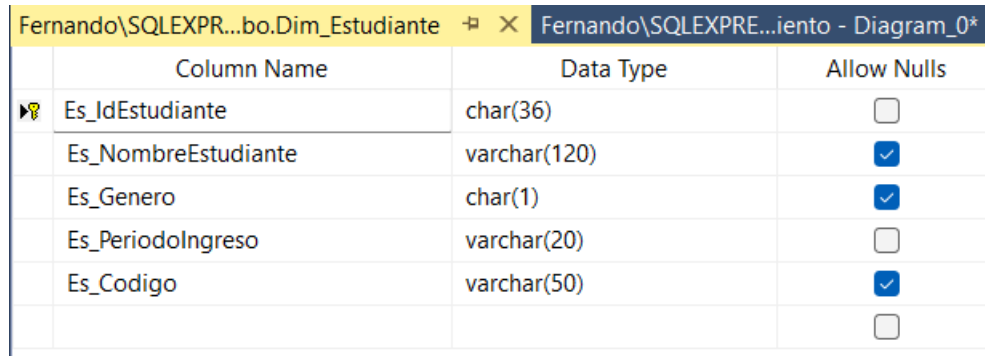


4.1.6.1. Construcción de las tablas y la base de datos

A continuación, se desarrolló la construcción de las tablas en el motor de la base de datos SQL Server.

Figura 25

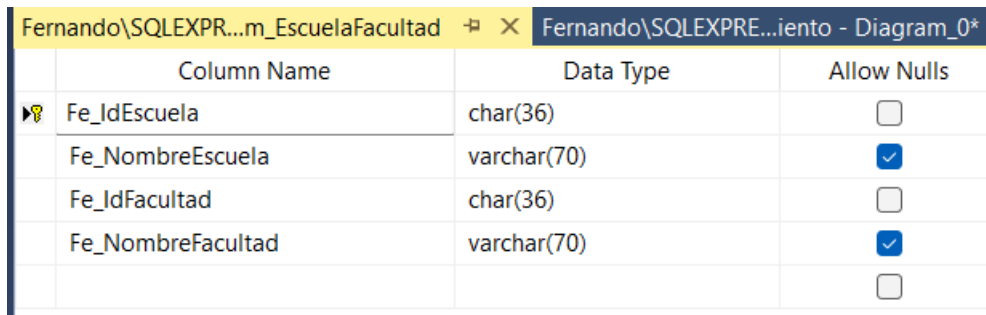
Tabla Dim_Estudiante



Column Name	Data Type	Allow Nulls
Es_IdEstudiante	char(36)	<input type="checkbox"/>
Es_NombreEstudiante	varchar(120)	<input checked="" type="checkbox"/>
Es_Genero	char(1)	<input checked="" type="checkbox"/>
Es_PeriodoIngreso	varchar(20)	<input type="checkbox"/>
Es_Codigo	varchar(50)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Figura 26

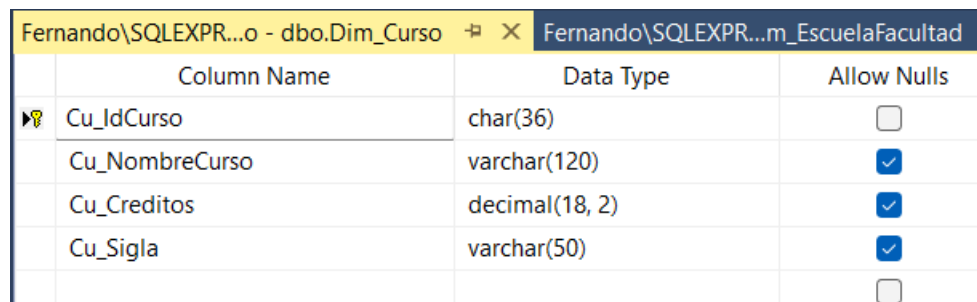
Tabla Dim_EscuelaFacultad



Column Name	Data Type	Allow Nulls
Fe_IdEscuela	char(36)	<input type="checkbox"/>
Fe_NombreEscuela	varchar(70)	<input checked="" type="checkbox"/>
Fe_IdFacultad	char(36)	<input type="checkbox"/>
Fe_NombreFacultad	varchar(70)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Figura 27

Tabla Dim_Curso



Column Name	Data Type	Allow Nulls
Cu_IdCurso	char(36)	<input type="checkbox"/>
Cu_NombreCurso	varchar(120)	<input checked="" type="checkbox"/>
Cu_Creditos	decimal(18, 2)	<input checked="" type="checkbox"/>
Cu_Sigla	varchar(50)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Figura 28

Tabla Dim_PeriodoAcademico

Column Name	Data Type	Allow Nulls
Pa_IdPeriodoAcademico	char(36)	<input type="checkbox"/>
Pa_AnioAcademico	varchar(50)	<input checked="" type="checkbox"/>
Pa_Semestre	varchar(50)	<input checked="" type="checkbox"/>
Pa_FechaInicio	date	<input checked="" type="checkbox"/>
Pa_FechaFin	date	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Figura 29

Tabla Dim_PlanEstudio

Column Name	Data Type	Allow Nulls
P_IdPlan	char(36)	<input type="checkbox"/>
P_PlanEstudio	varchar(50)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

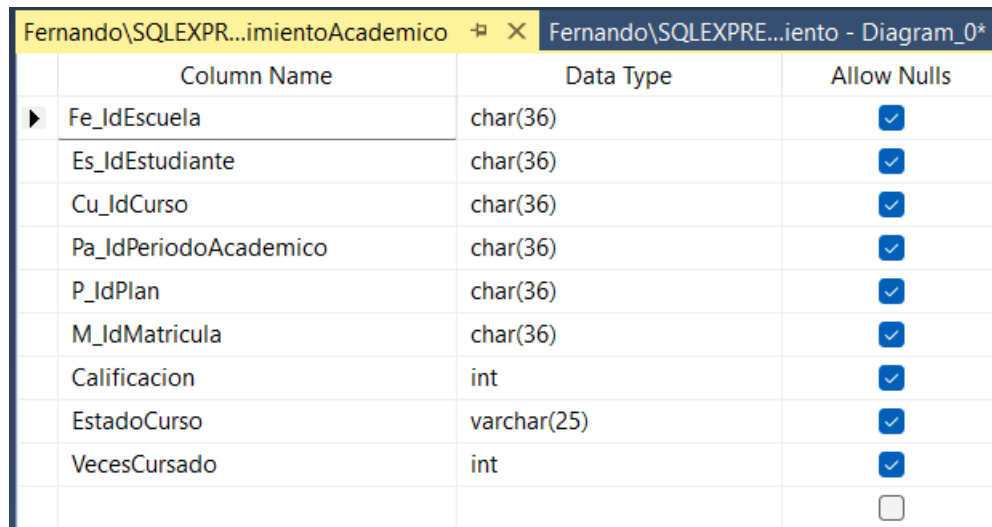
Figura 30

Tabla Dim_Matricula

Column Name	Data Type	Allow Nulls
M_IdMatricula	char(36)	<input type="checkbox"/>
M_SeccionMatricula	varchar(50)	<input checked="" type="checkbox"/>
M_Vacantes	int	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Figura 31

Tabla Fact_SeguimientoAcademico

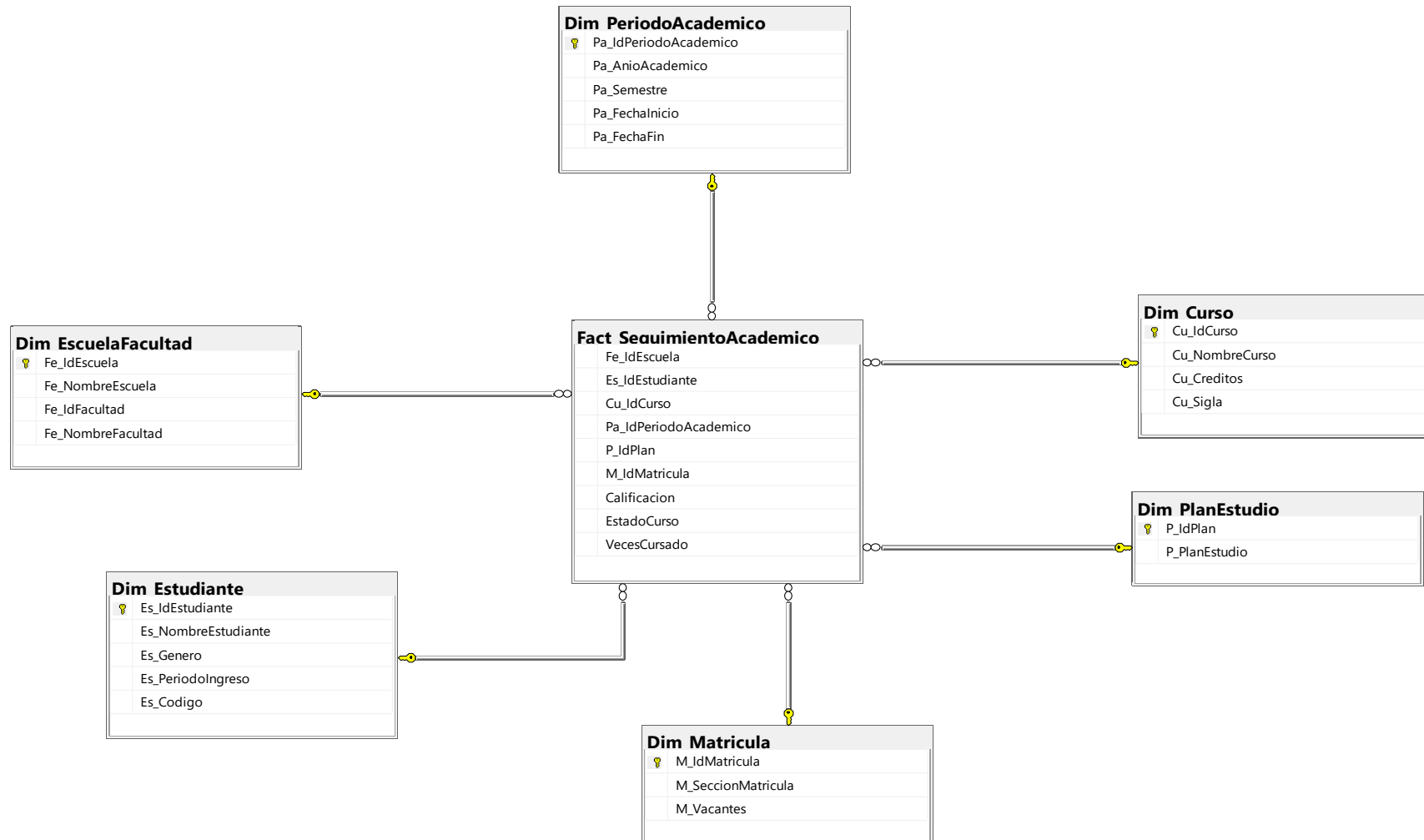


Column Name	Data Type	Allow Nulls
Fe_IdEscuela	char(36)	<input checked="" type="checkbox"/>
Es_IdEstudiante	char(36)	<input checked="" type="checkbox"/>
Cu_IdCurso	char(36)	<input checked="" type="checkbox"/>
Pa_IdPeriodoAcademico	char(36)	<input checked="" type="checkbox"/>
P_IdPlan	char(36)	<input checked="" type="checkbox"/>
M_IdMatricula	char(36)	<input checked="" type="checkbox"/>
Calificacion	int	<input checked="" type="checkbox"/>
EstadoCurso	varchar(25)	<input checked="" type="checkbox"/>
VecesCursado	int	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Construidas las dimensiones y la tabla de hechos del Datamart, a continuación, se mostrará el diagrama total de la base datos en SQL Server.

Figura 32

Modelo físico del Datamart en SQL Server



4.1.7. Diseño e implementación del sub sistema ETL

4.1.7.1. Fases del proceso ETL

La metodología Kimball, nos indica que para realizar la construcción del Datamart, se debe poblar la base de datos, como ya se tiene el modelo físico ya construido en el motor de la base de datos del SQL Server, siguiendo las fases del proceso ETL, utilizando la herramienta integración de Servicios del SQL Server.

A continuación, se detallan las fases del proceso ETL:

Paso 1: Limpiar tablas. En este paso se borraron todos los registros de las tablas, relaciones, tabla de hechos y dimensiones, para asegurar que no existan algún dato registrado.

Paso 2: Poblamiento de la dimensión EscuelaFacultad. Se carga la dimensión con la ejecución de una sentencia SQL, que se detallara posteriormente.

Paso 3: Poblamiento de la dimensión Estudiante. Se carga la dimensión con la ejecución de una sentencia SQL, que se detallara posteriormente.

Paso 4: Poblamiento de la dimensión Curso. Se carga la dimensión con la ejecución de una sentencia SQL, que se detallara posteriormente.

Paso 5: Poblamiento de la dimensión Periodo Académico. Se carga la dimensión con la ejecución de una sentencia SQL, que se detallara posteriormente.

Paso 6: Poblamiento de la dimensión Plan Estudio. Se carga la dimensión con la ejecución de una sentencia SQL, que se detallara posteriormente.

Paso 7: Poblamiento de la dimensión Matricula. Se carga la dimensión con la ejecución de una sentencia SQL, que se detallara posteriormente.

Paso 8: Poblamiento de la tabla hecho.

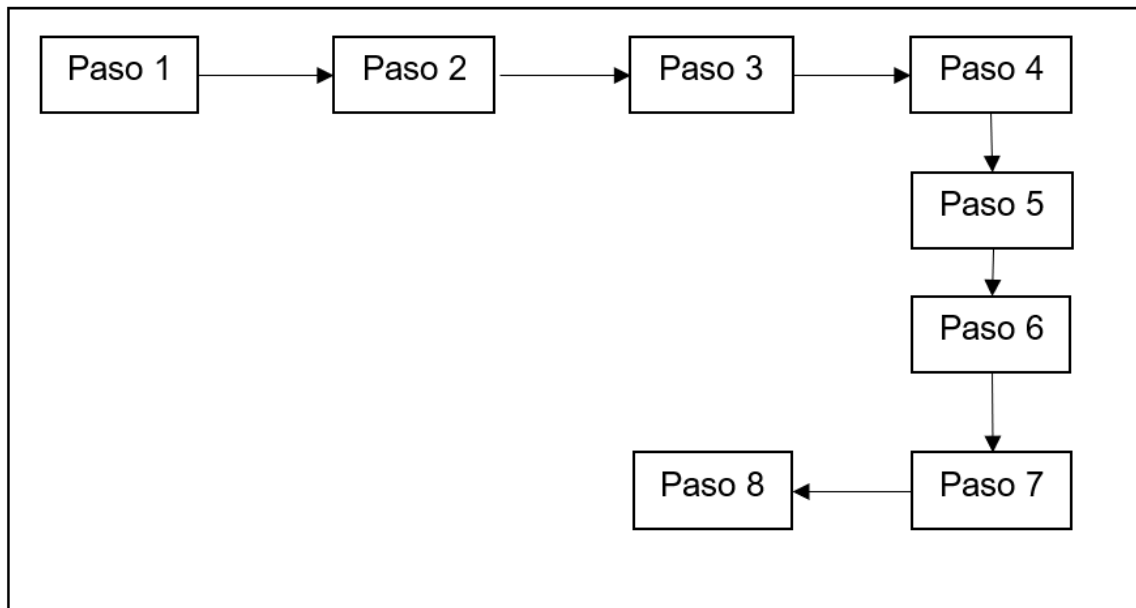
a. Calificación. Se realiza el proceso de carga de la tabla de hecho Calificación.

b. **VecesCursado.** Se realiza el proceso de carga de la tabla de hecho VecesCursado.

c. **EstadoCurso.** Se realiza el proceso de carga de la tabla de hecho EstadoCurso.

Figura 33

Diagrama de pasos del proceso ETL



- Restricciones de precedencia:

a. Se procede con la limpieza de las dimensiones (Paso 1).

b. Se realizo el poblamiento de la dimensión Escuela Facultad (Paso 2), se procederá con este paso cuando este completo el paso 1.

c. Se realizo el poblamiento de la dimensión Estudiante (Paso 3), se procederá con este paso cuando este completo el paso 1.

d. Se realizo el poblamiento de la dimensión Curso (Paso 4), se procederá con este paso cuando este completo el paso 1.

e. Se realizo el poblamiento de la dimensión Periodo Académico (Paso 5), se procederá con este paso cuando este completo el paso 1.

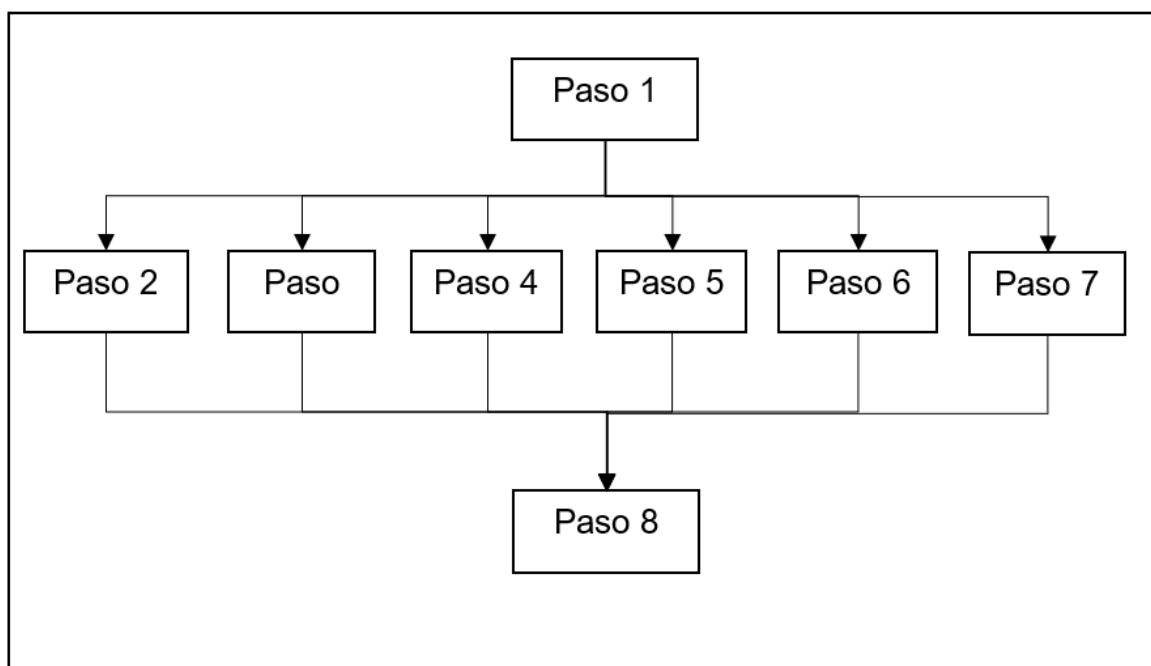
f. Se realizo el poblamiento de la dimensión Plan Estudio (Paso 6), se procederá con este paso cuando este completo el paso 1.

g. Se realizo el poblamiento de la dimensión Matricula (Paso 7), se procederá con este paso cuando este completo el paso 1.

h. Se realizo el poblamiento de la tabla de hecho Seguimiento Académico (Paso 8), se procederá con este paso cuando estén completos los pasos 2,3,4,5,6,7.

Figura 34

Diagrama de pasos con restricciones de precedencia



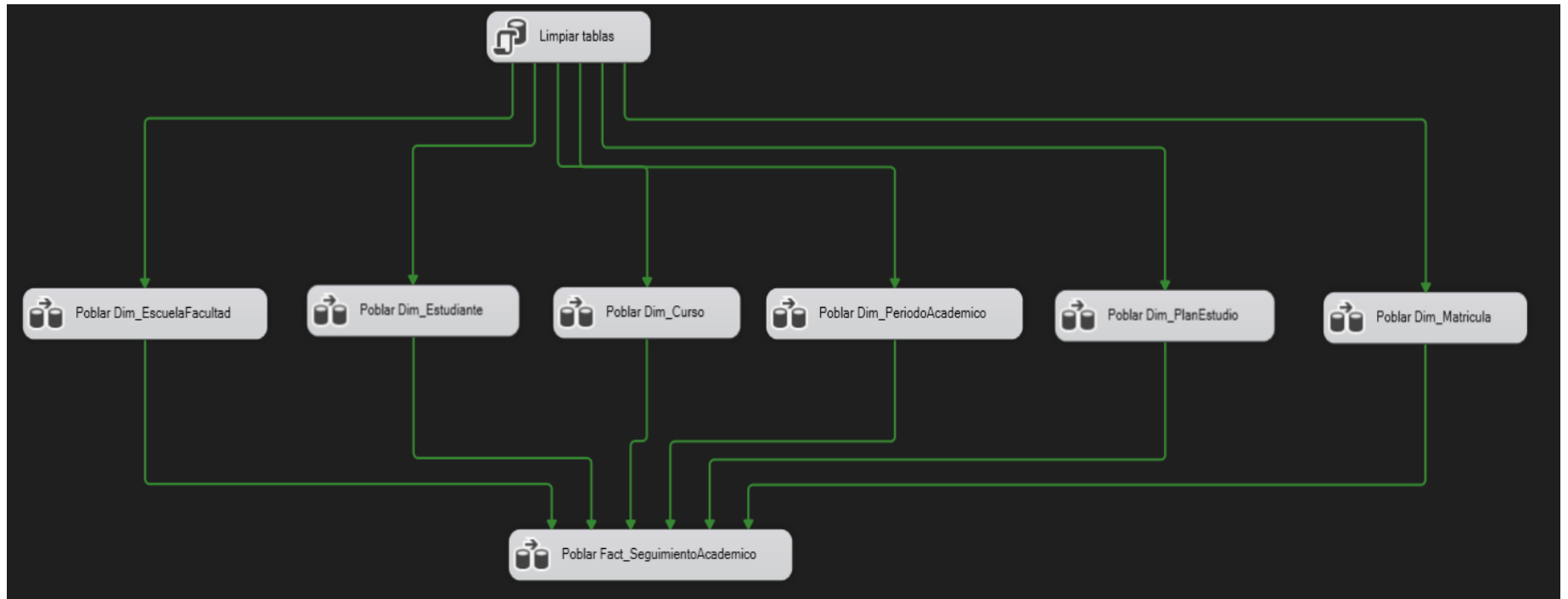
4.1.7.2. Pasos del proceso ETL

Carga inicial

Se comienza a llenar las dimensiones descritas, con el Visual Studio.

Figura 35

Proceso ETL de la carga inicial



1. Limpiando dimensiones

Figura 36

Comando SQL para la limpieza de las tablas del Datamart

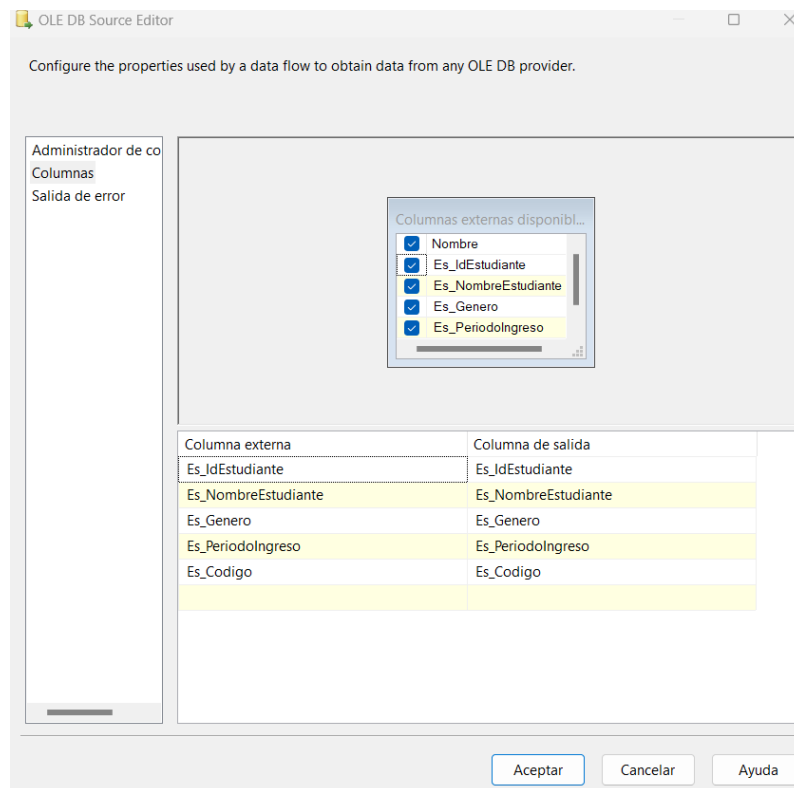
```
delete from Fact_SeguimientoAcademico;  
delete from Dim_Estudiante;  
delete from Dim_EscuelaFacultad;  
delete from Dim_Curso;  
delete from Dim_PeriodoAcademico;  
delete from Dim_PlanEstudio;  
delete from Dim_Matricula;
```

2. Poblamiento de la dimensión Estudiante

A. Origen de datos

Figura 37

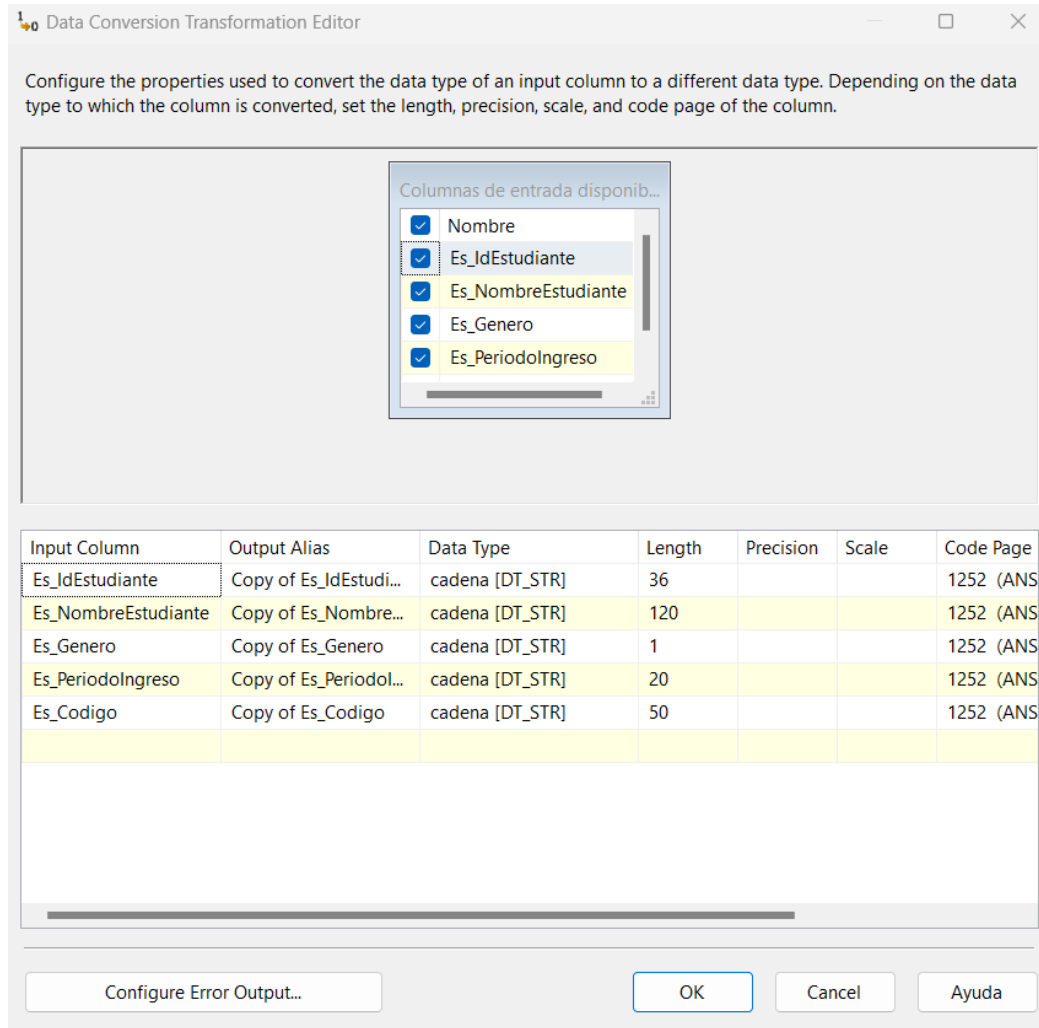
Origen de datos para la dimensión Estudiante



B. Conversión de datos

Figura 38

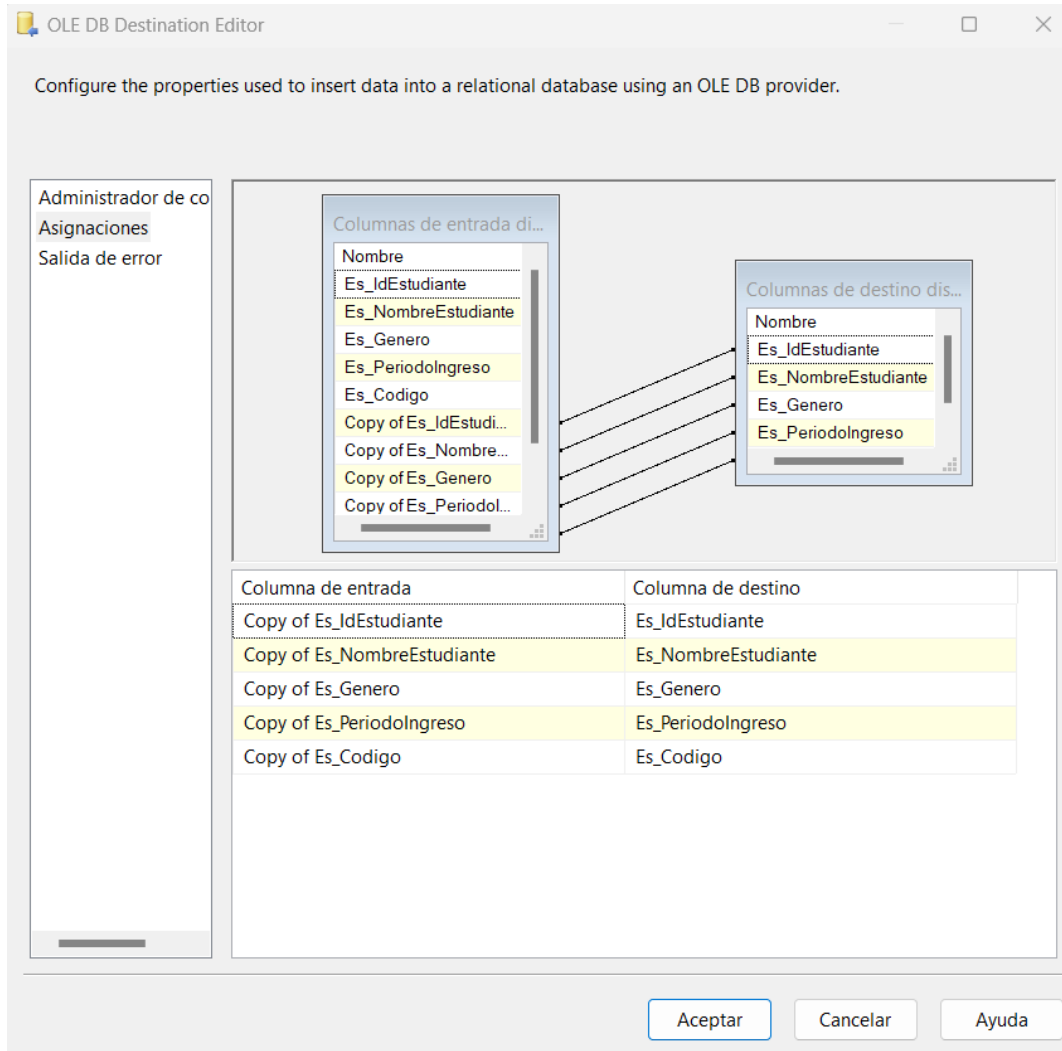
Conversión de datos para la dimensión Estudiante



C. Poblar datos de la dimensión Estudiante

Figura 39

Carga de datos para la dimensión Estudiante

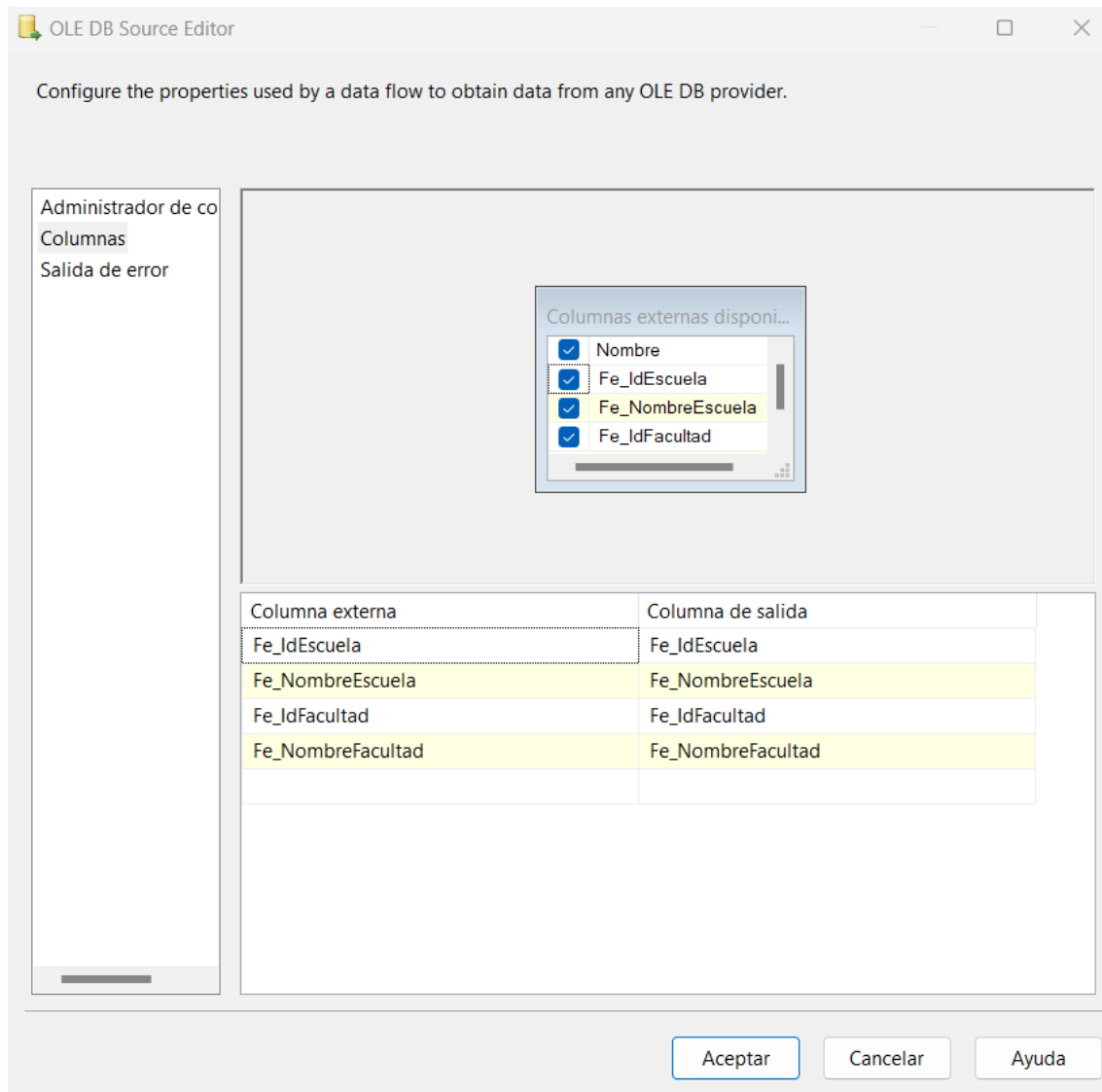


3. Poblamiento de la dimensión Escuela Facultad

A. Origen de datos

Figura 40

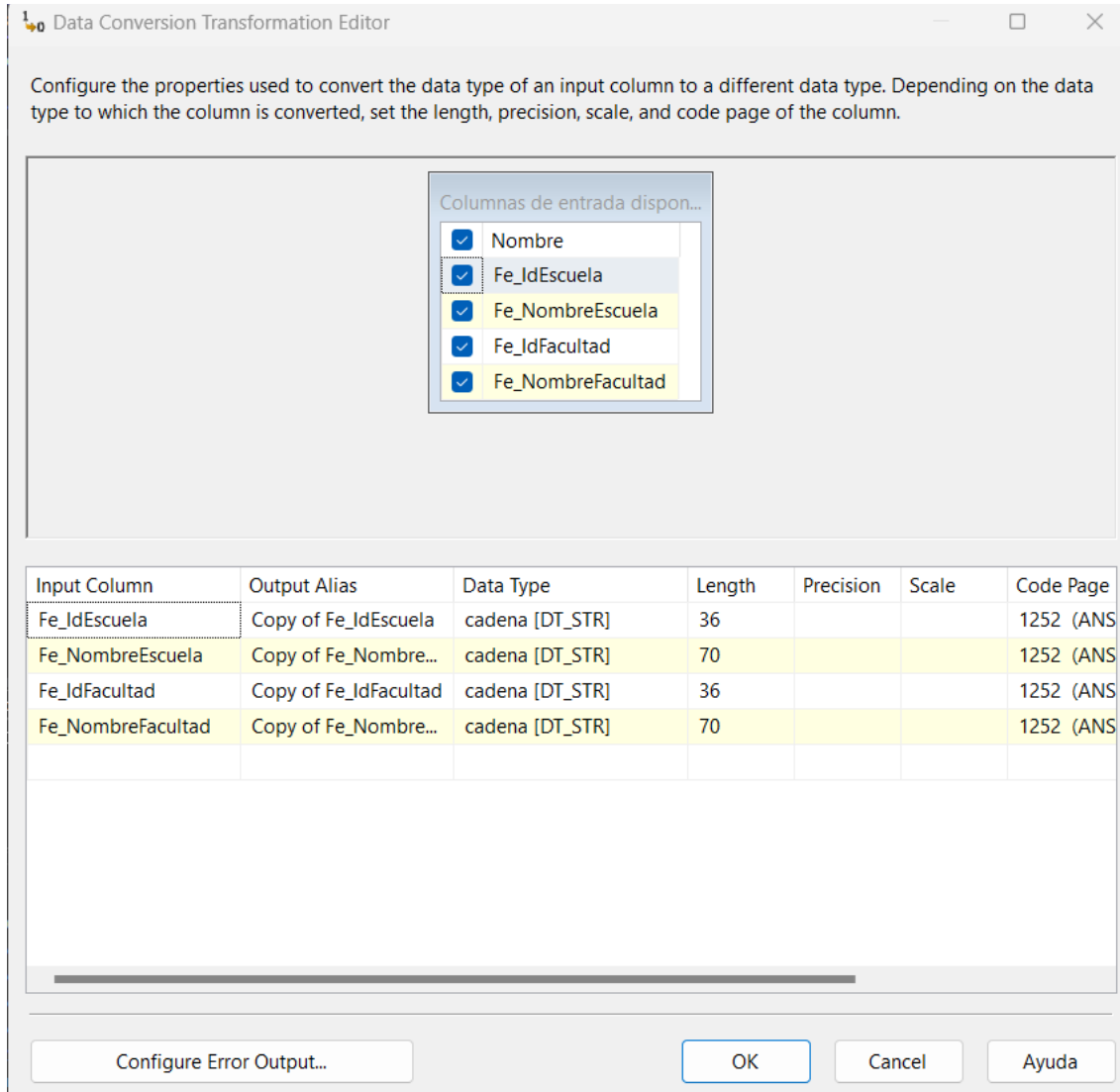
Origen de datos para la dimensión Escuela Facultad



B. Conversión de datos

Figura 41

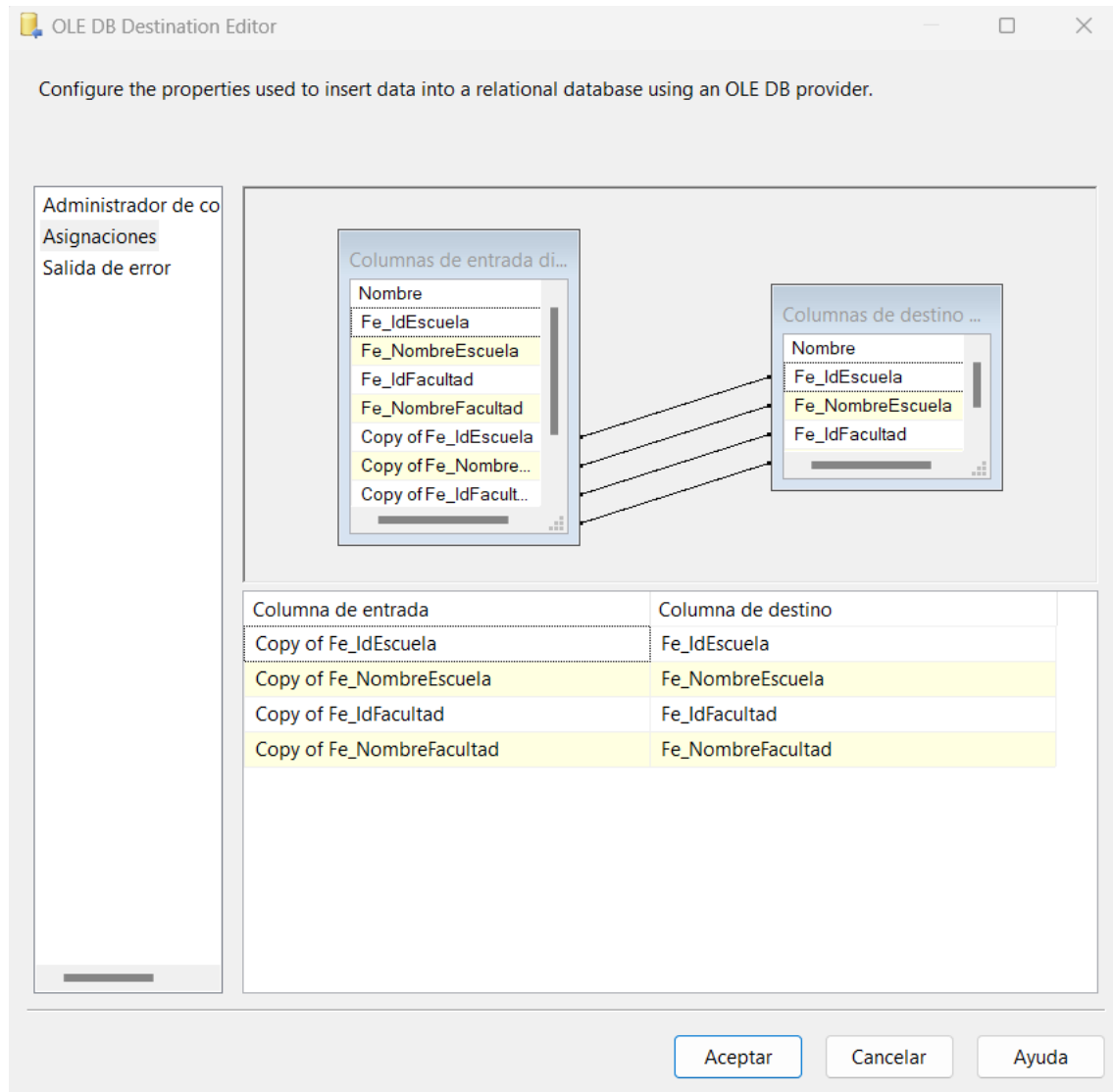
Conversión de datos para la dimensión Escuela Facultad



C. Poblar datos de la dimensión Escuela Facultad

Figura 42

Carga de datos para la dimensión Escuela Facultad

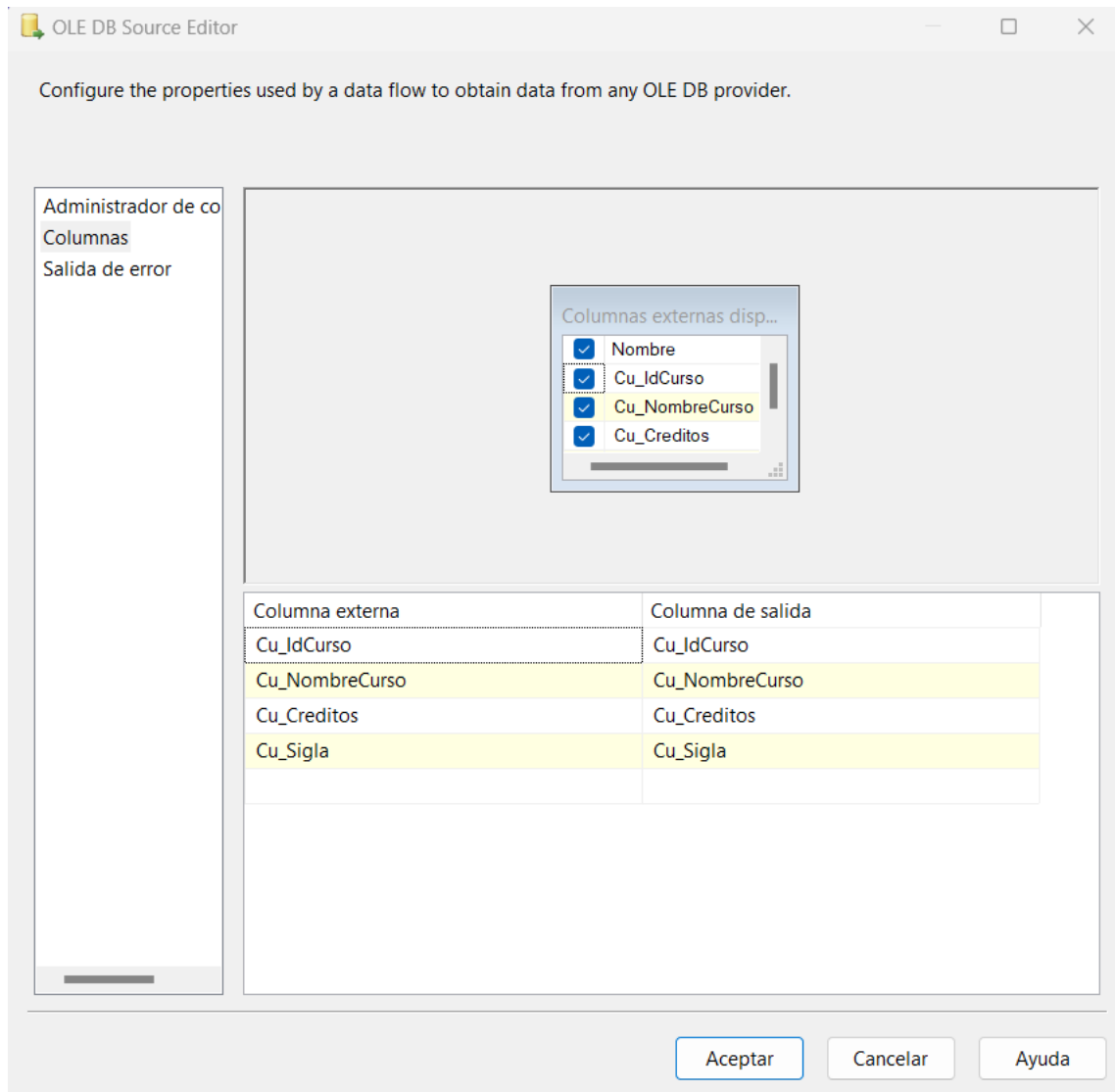


4. Poblamiento de la dimensión Curso

A. Origen de datos

Figura 43

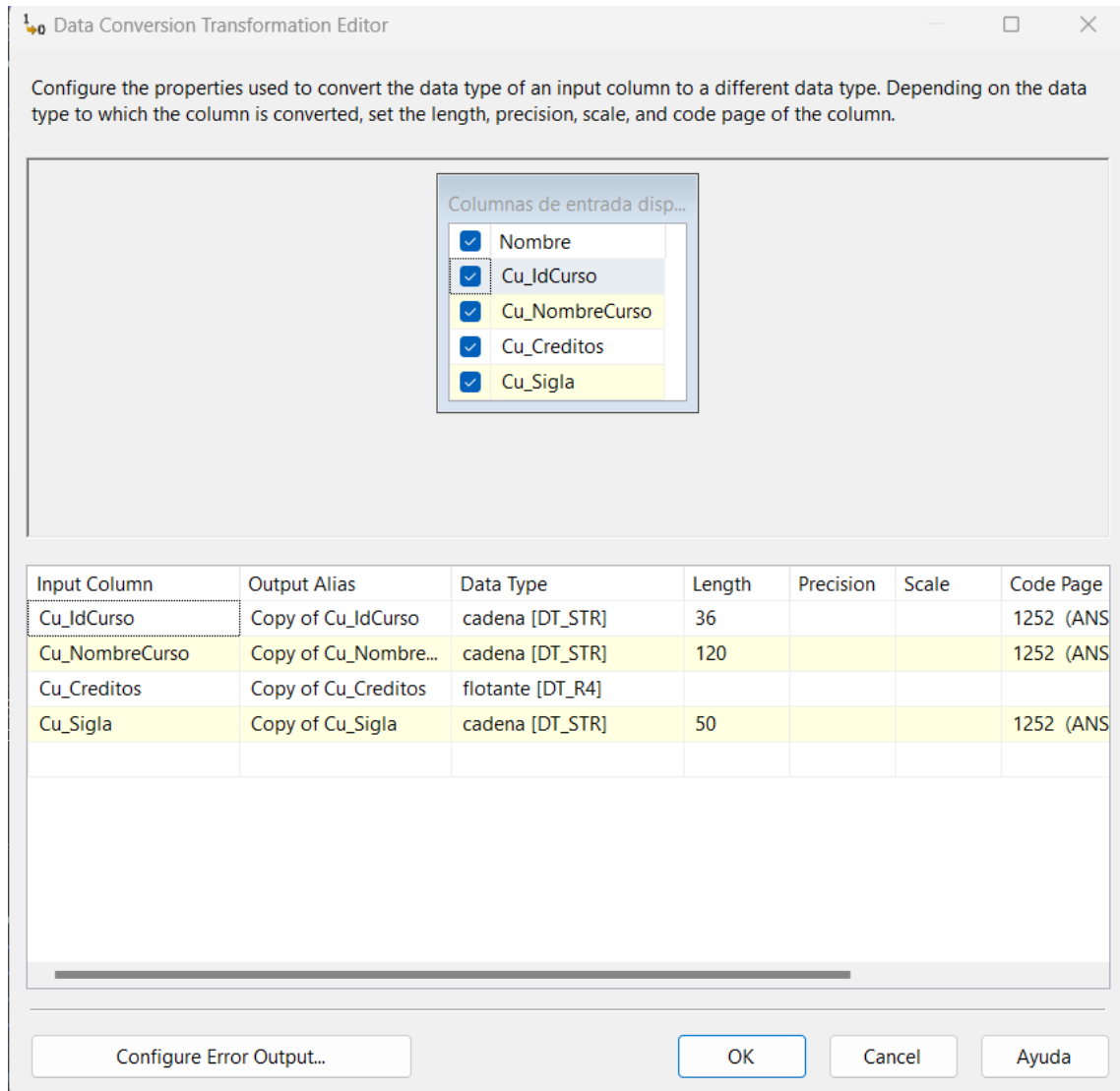
Origen de datos para la dimensión Curso



B. Conversión de datos

Figura 44

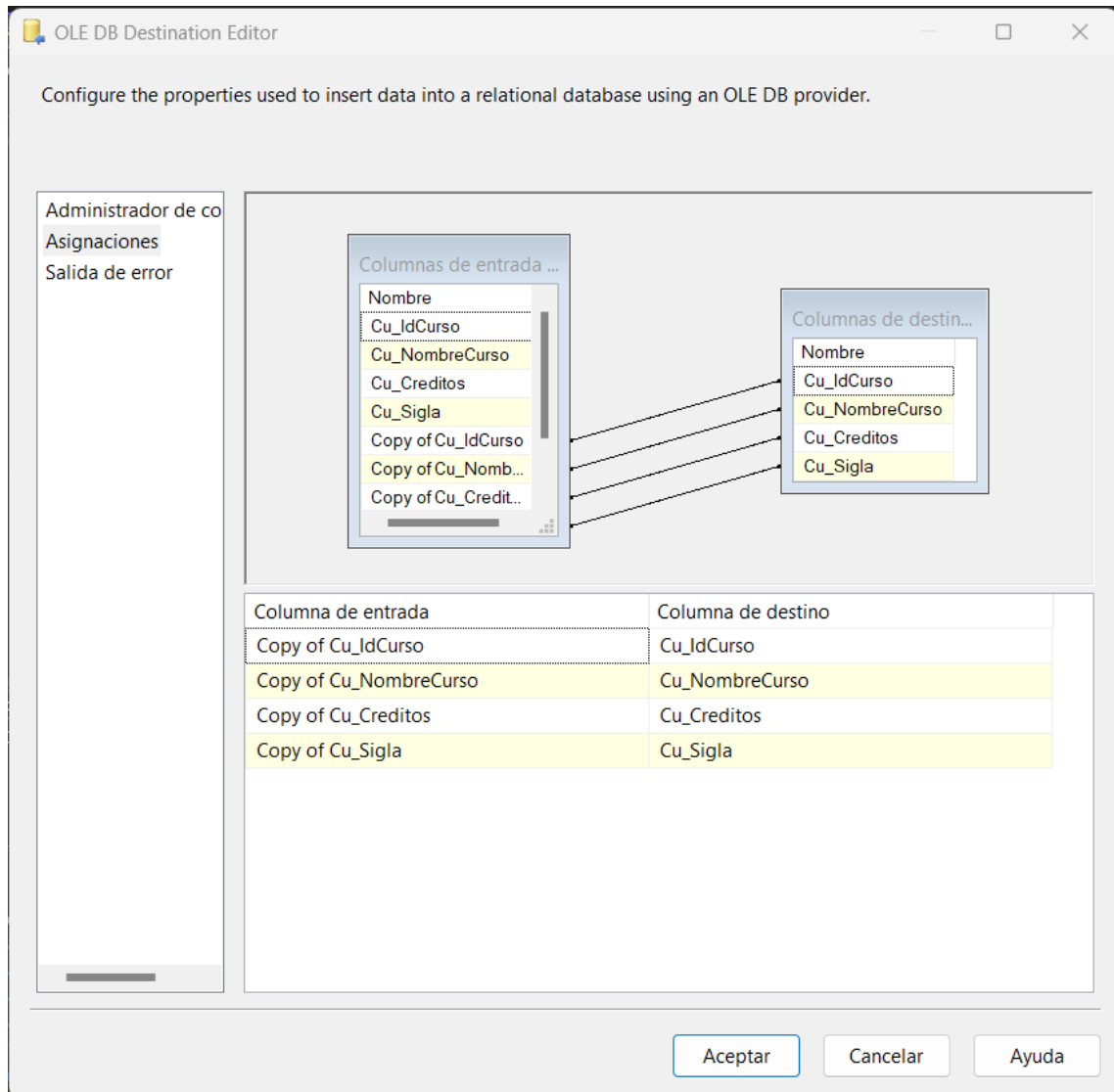
Conversión de datos para la dimensión Curso



C. Poblar datos de la dimensión Curso

Figura 45

Carga de datos para la dimensión Curso

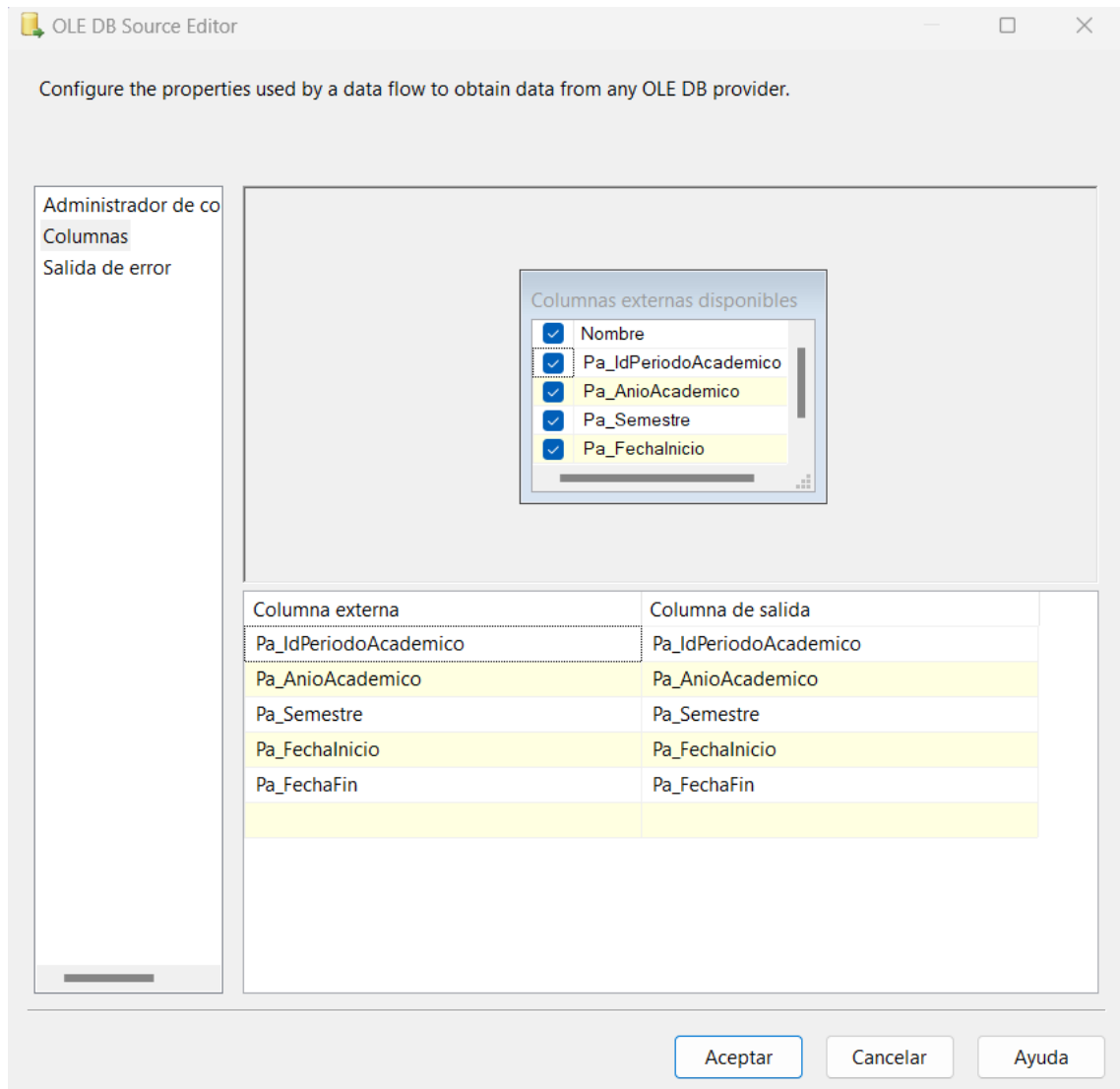


5. Poblamiento de la dimensión Período Académico

A. Origen de datos

Figura 46

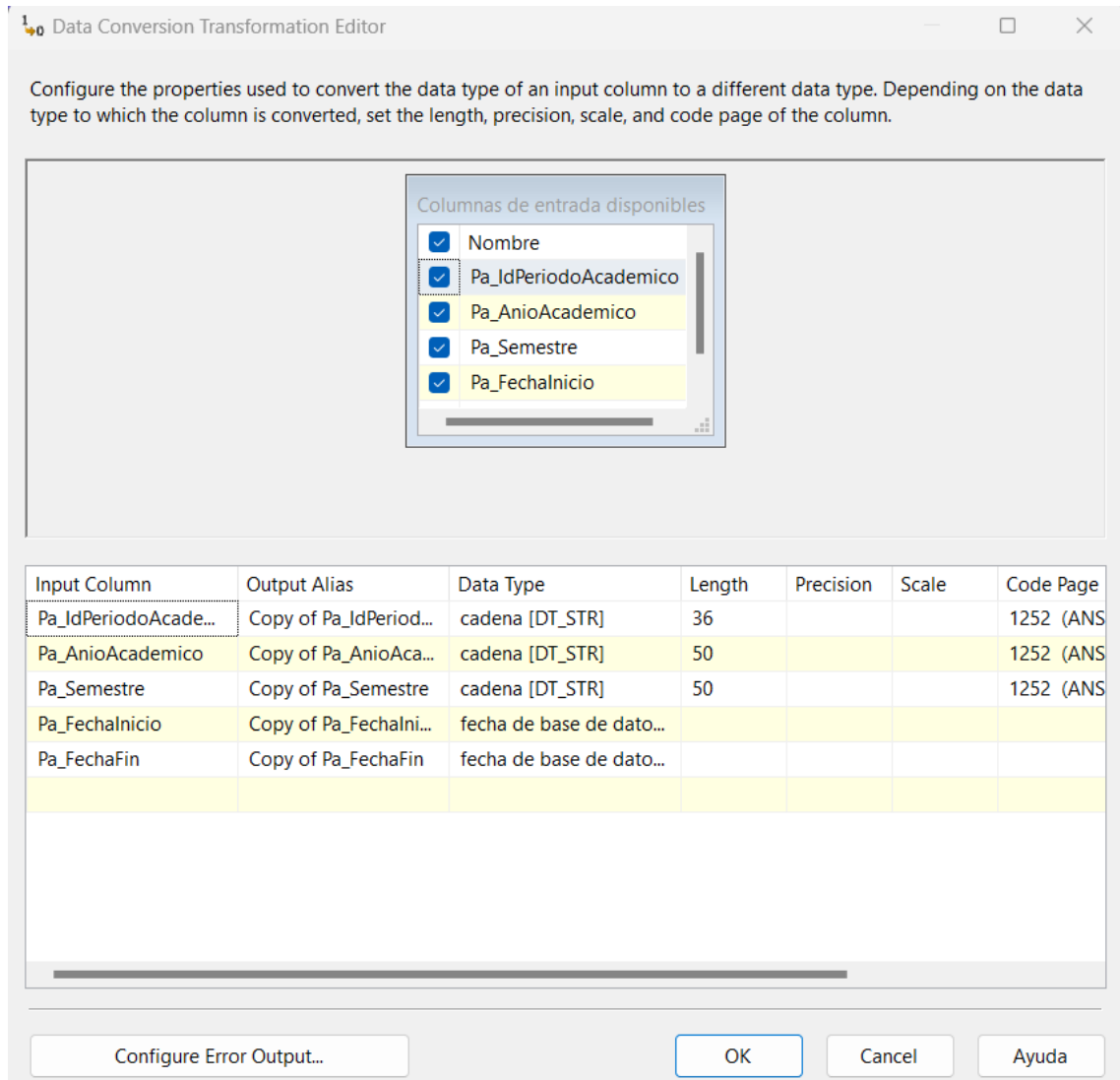
Origen de datos para la dimensión Período Académico



B. Conversión de datos

Figura 47

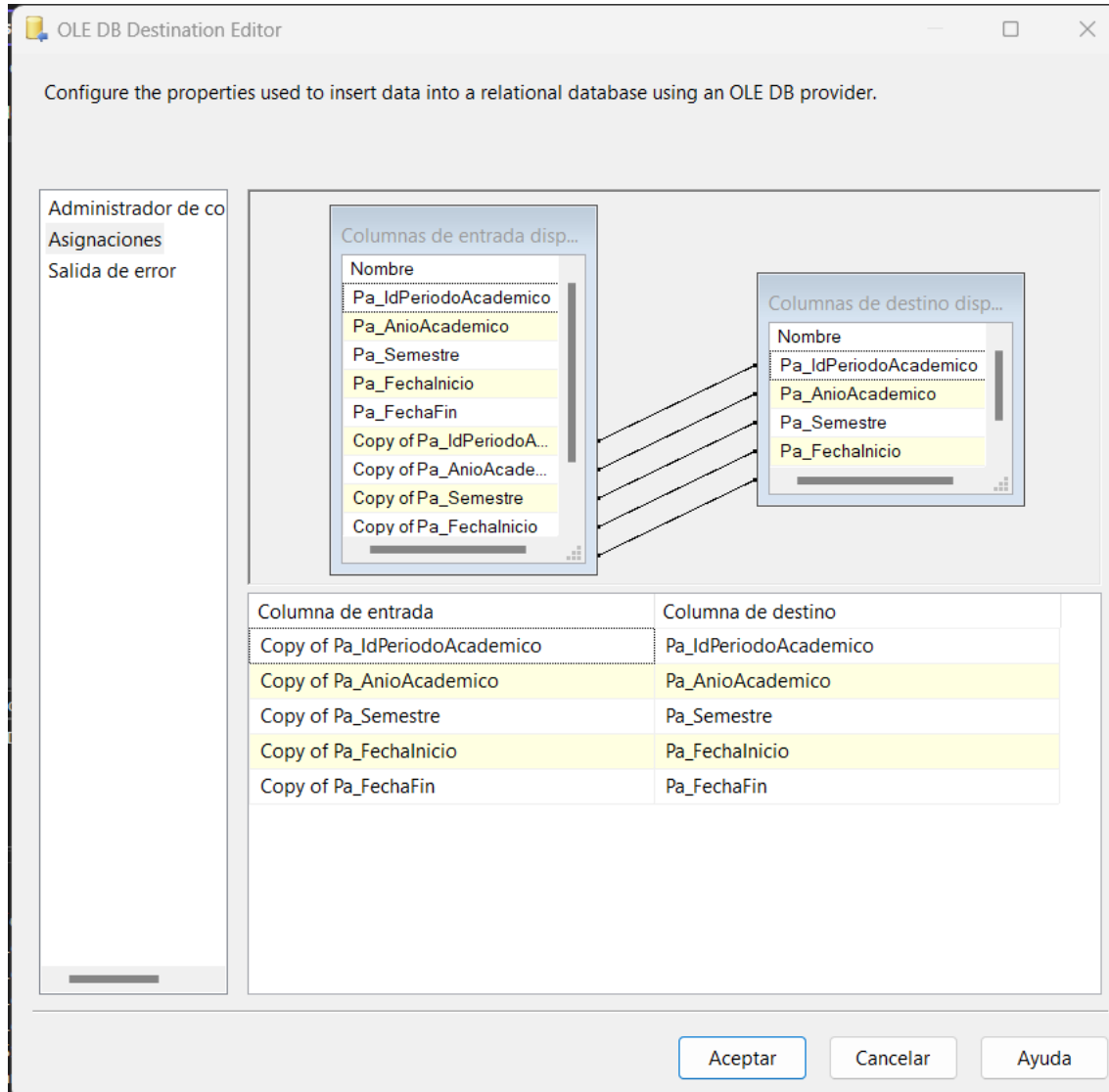
Conversión de datos para la dimensión Periodo Académico



C. Poblar datos de la dimensión Periodo Académico

Figura 48

Carga de datos para la dimensión Periodo Académico

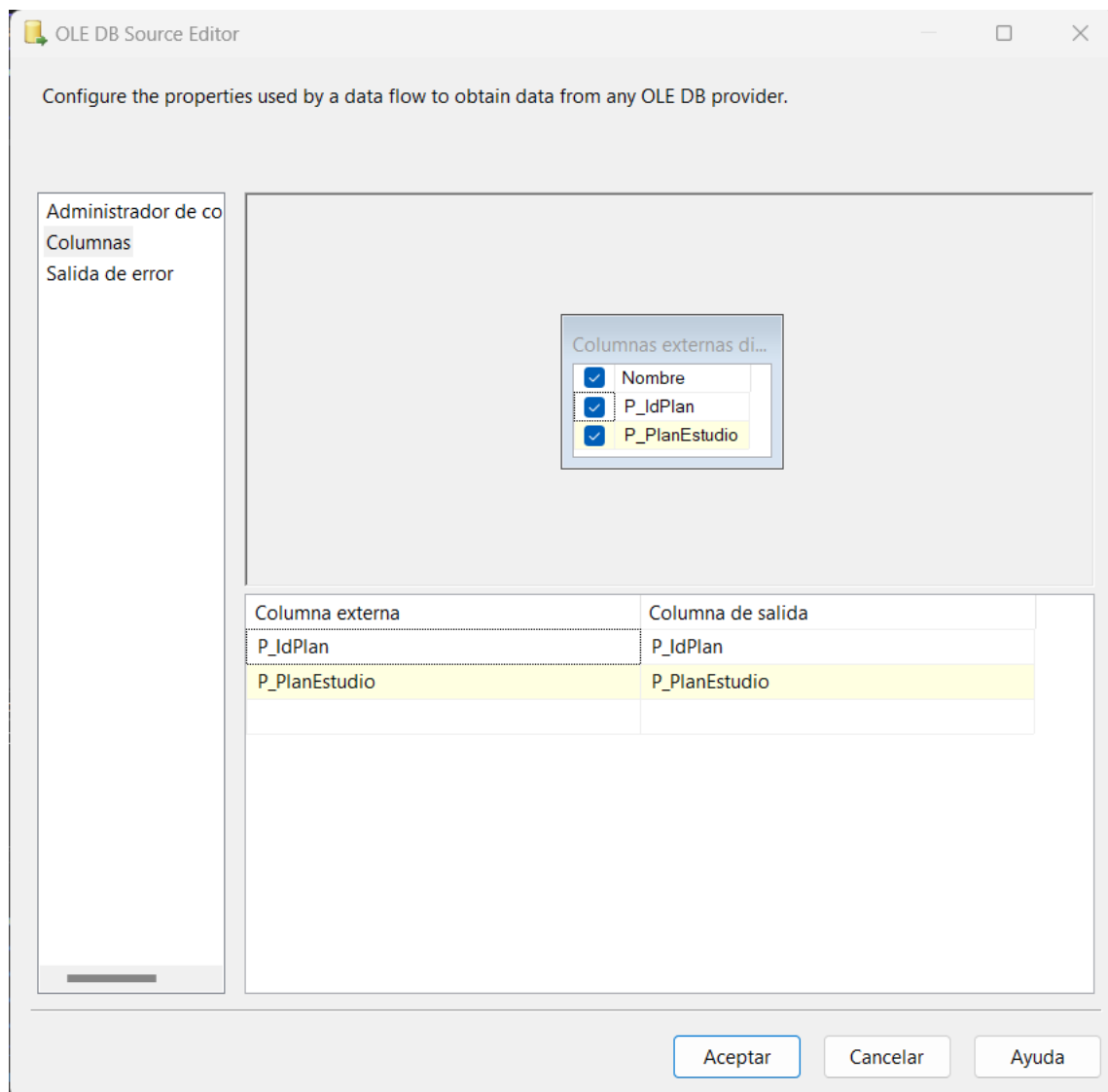


6. Poblamiento de la dimensión Plan Estudio

A. Origen de datos

Figura 49

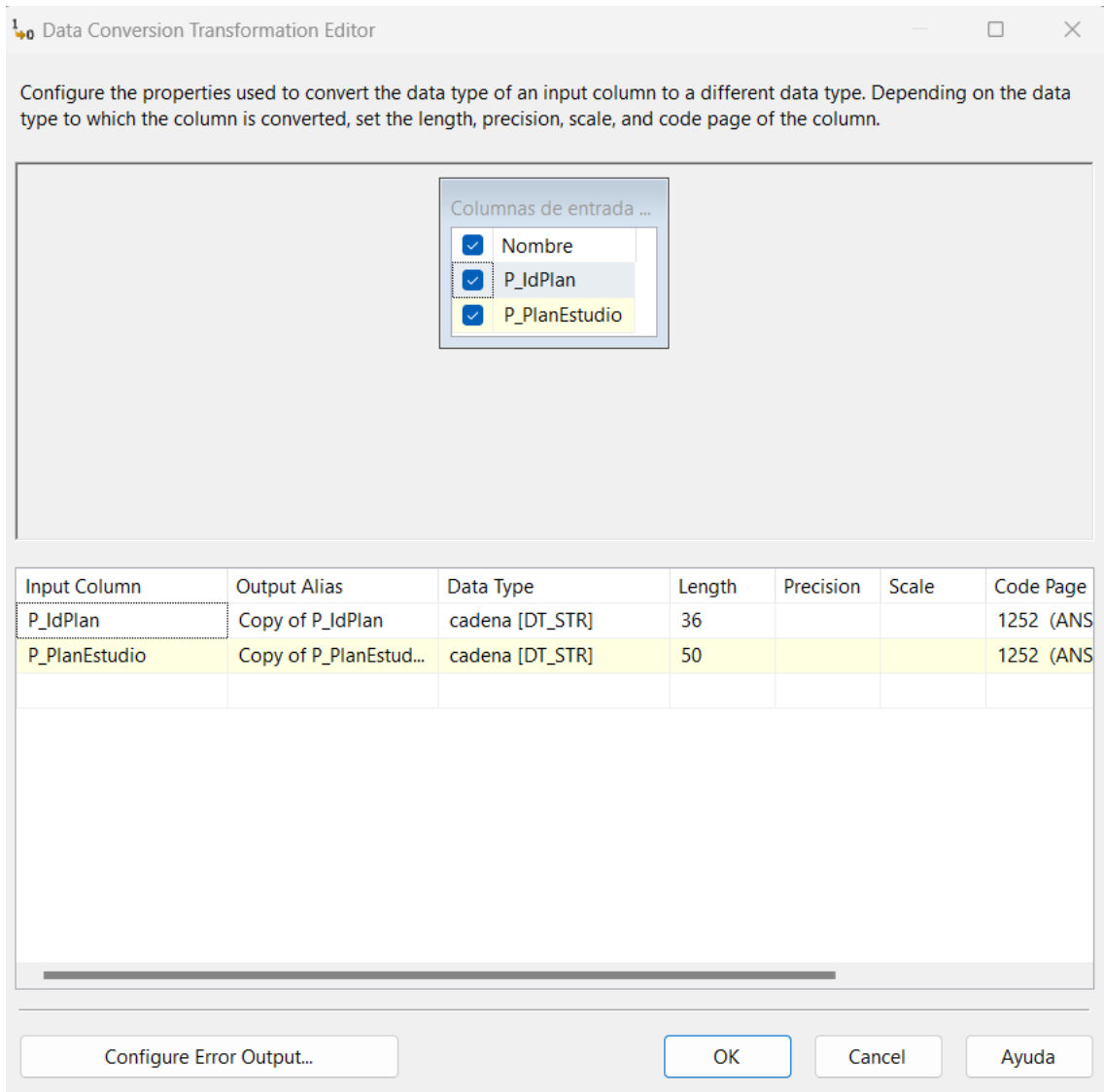
Origen de datos para la dimensión Plan Estudio



B. Conversión de datos

Figura 50

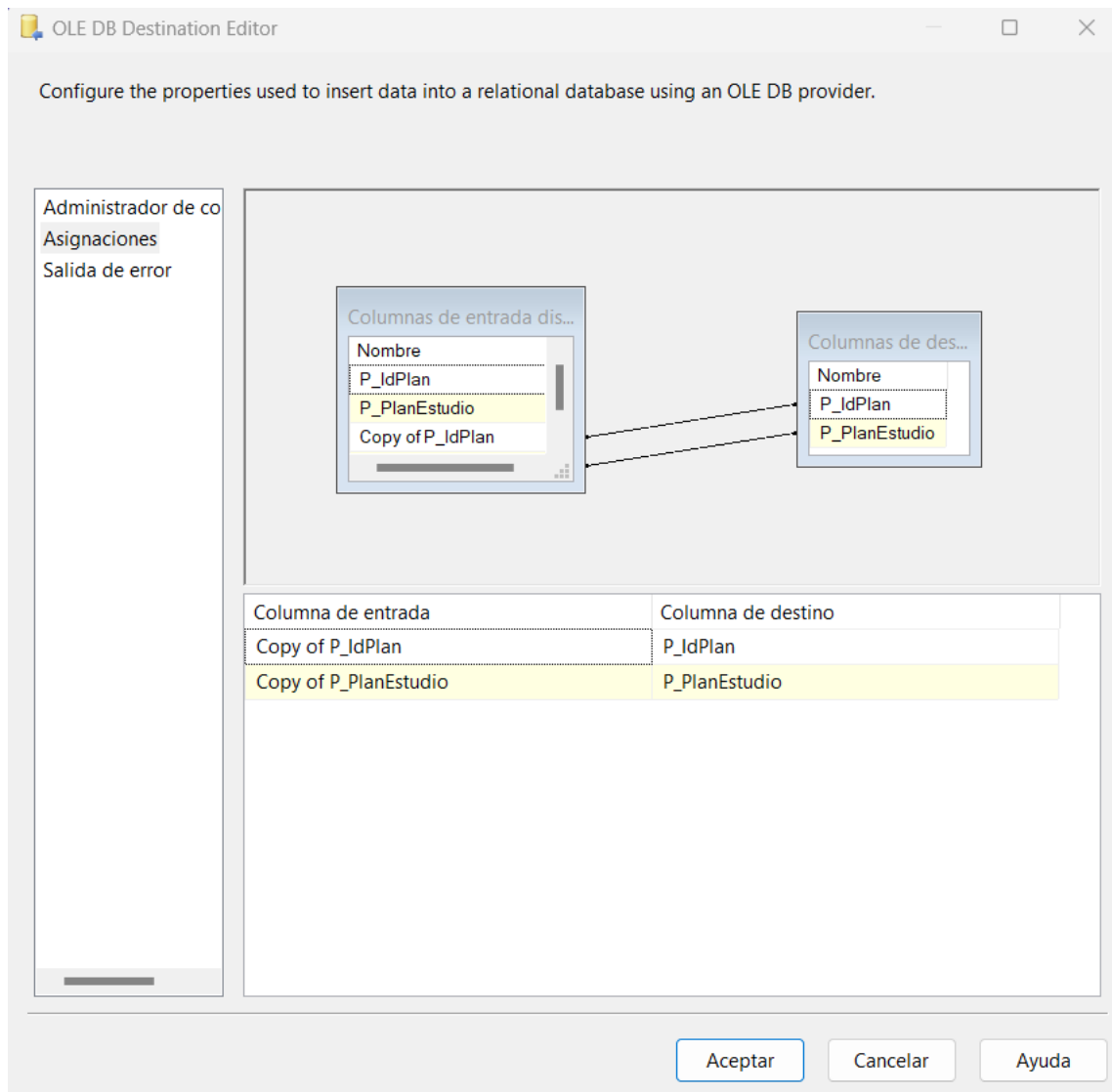
Conversión de datos para la dimensión Plan Estudio



C. Poblar datos de la dimensión Plan Estudio

Figura 51

Carga de datos para la dimensión Plan Estudio

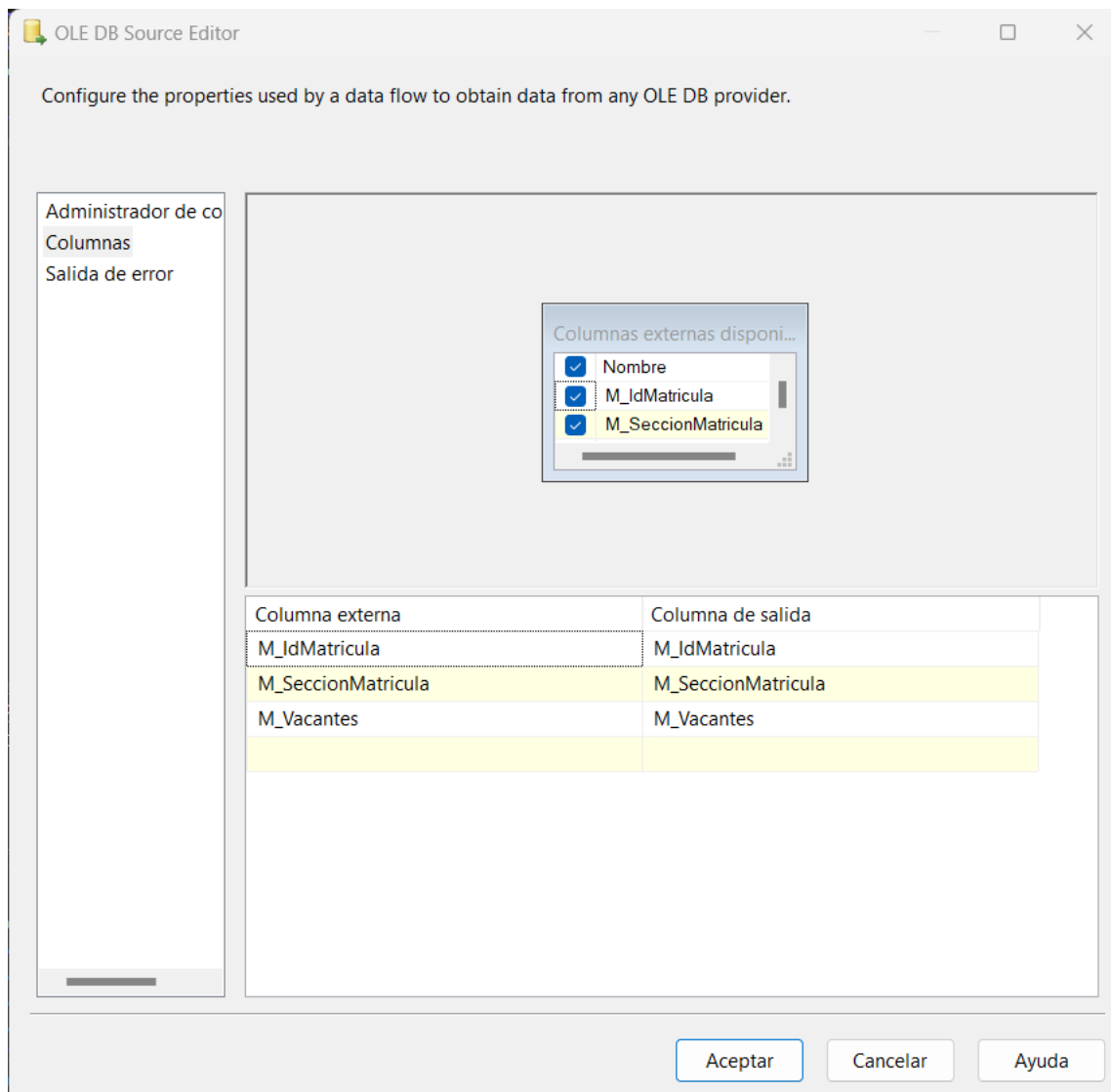


7. Poblamiento de la dimensión Matricula

A. Origen de datos

Figura 52

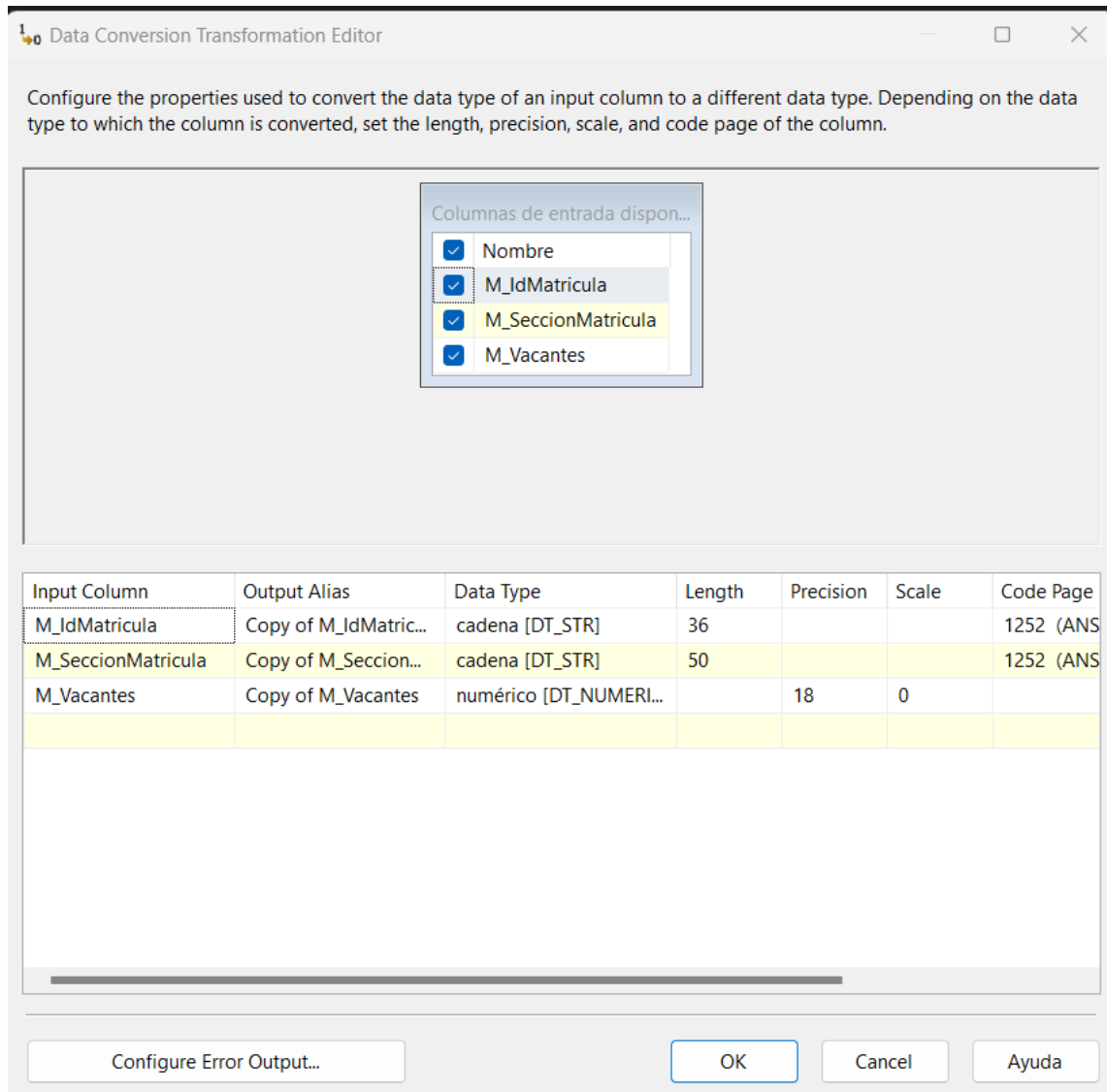
Origen de datos para la dimensión Matricula



B. Conversión de datos

Figura 53

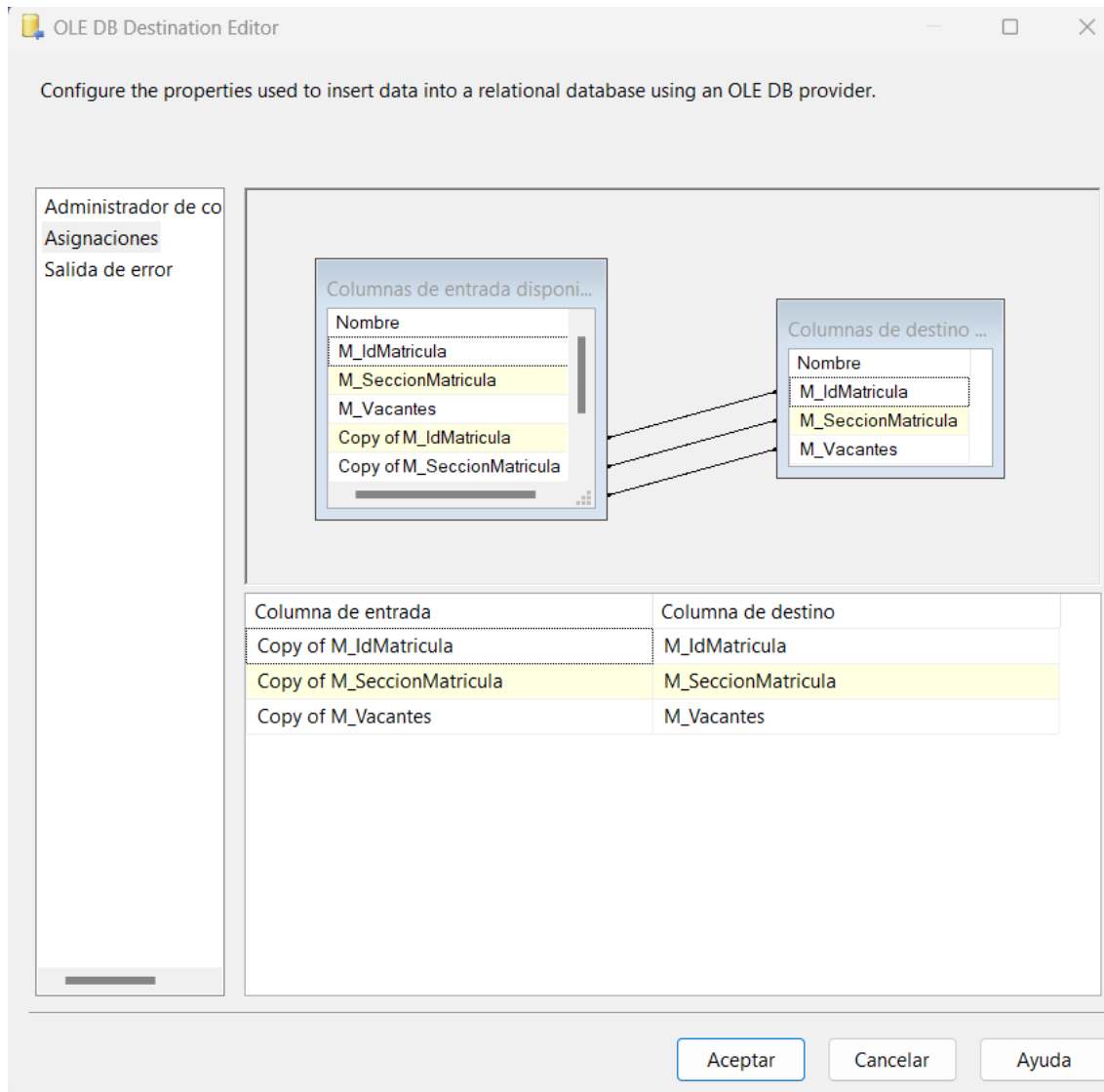
Conversión de datos para la dimensión Matricula



C. Poblar datos de la dimensión Matricula

Figura 54

Carga de datos para la dimensión Matricula

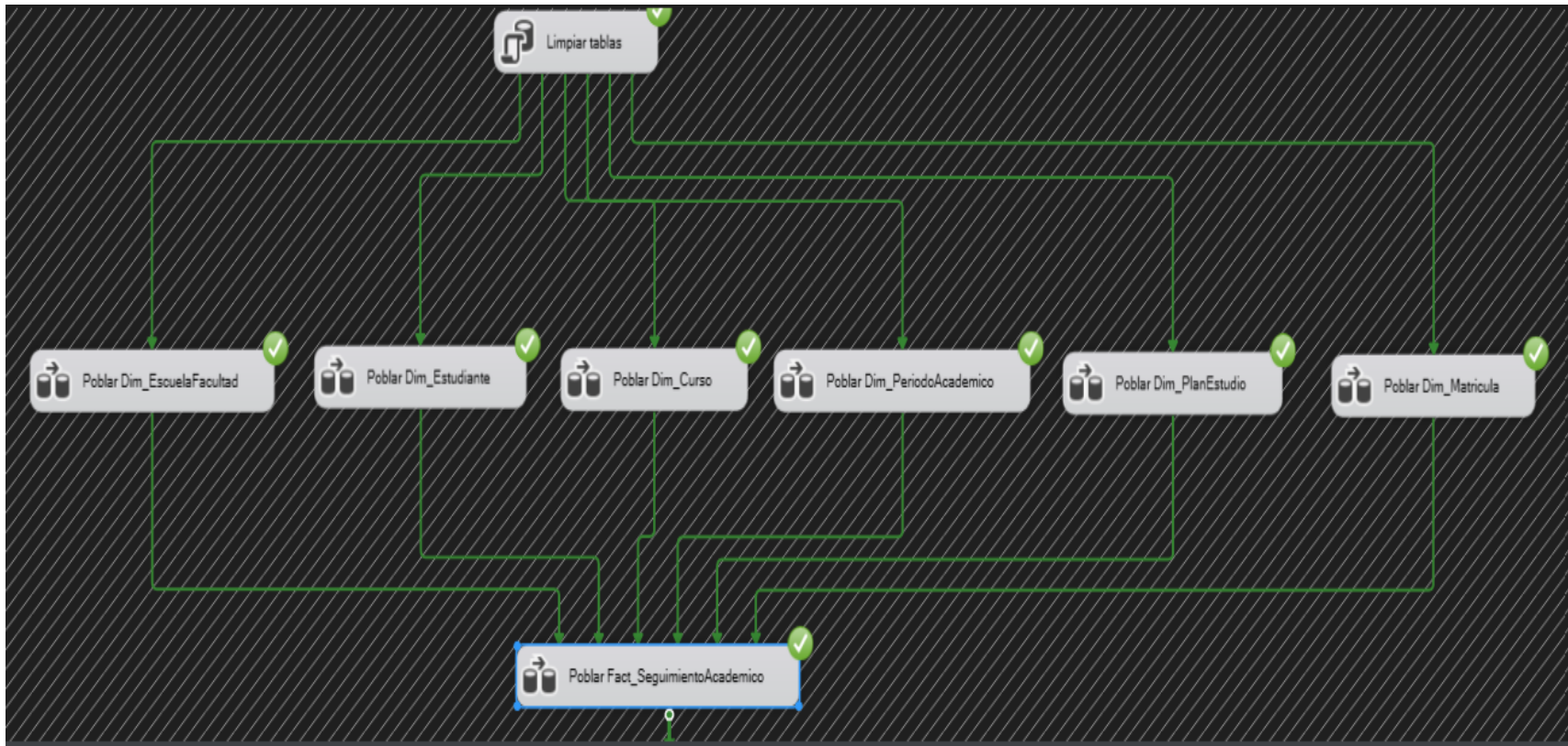


8. Poblamiento de la tabla de hechos Fact_SeguimientoAcademico

```
select fad.Fe_IdEscuela, esd.Es_IdEstudiante, cud.Cu_IdCurso,
pad.Pa_IdPeriodoAcademico,
pd.P_IdPlan, md.M_IdMatricula,
iah.Grade as Calificacion,
case
    when iah.Approved=0 then 'Desaprobado'
    when iah.Approved=1 then 'Aprobado'
End EstadoCurso,
iah.Try as VecesCursado
from db_sigau.[unsch.sigau.db].enrollment_faculties efa
inner join db_sigau.[unsch.sigau.db].generals_careers gca on efa.Id=gca.FacultyId
inner join db_sigau.[unsch.sigau.db].generals_students gstu on gstu.CareerId=gca.Id
inner join db_sigau.[unsch.sigau.db].intranet_academichistories iah on gstu.Id=iah.StudentId
inner join db_sigau.[unsch.sigau.db].enrollment_courses ecu on ecu.Id=iah.CourseId
inner join db_sigau.[unsch.sigau.db].enrollment_terms ete on ete.Id=iah.TermId
inner join Datamart_Seguimiento.dbo.Dim_Estudiante esd on esd.Es_IdEstudiante=gstu.Id
inner join Datamart_Seguimiento.dbo.Dim_Curso cud on cud.Cu_IdCurso=ecu.Id
inner join Datamart_Seguimiento.dbo.Dim_EscuelaFacultad fad on fad.Fe_IdEscuela=gca.Id
inner join Datamart_Seguimiento.dbo.Dim_PeriodoAcademico pad on pad.Pa_IdPeriodoAcademico=ete.Id
inner join Datamart_Seguimiento.dbo.Dim_PlanEstudio pd on pd.P_IdPlan=gstu.CurriculumId
inner join Datamart_Seguimiento.dbo.Dim_Matricula md on md.M_IdMatricula=iah.SectionId
where iah.Type=1 and iah.Withdraw=0;
```

Figura 55

Proceso ETL completado para el Datamart



4.1.8. Implementación del cubo multidimensional

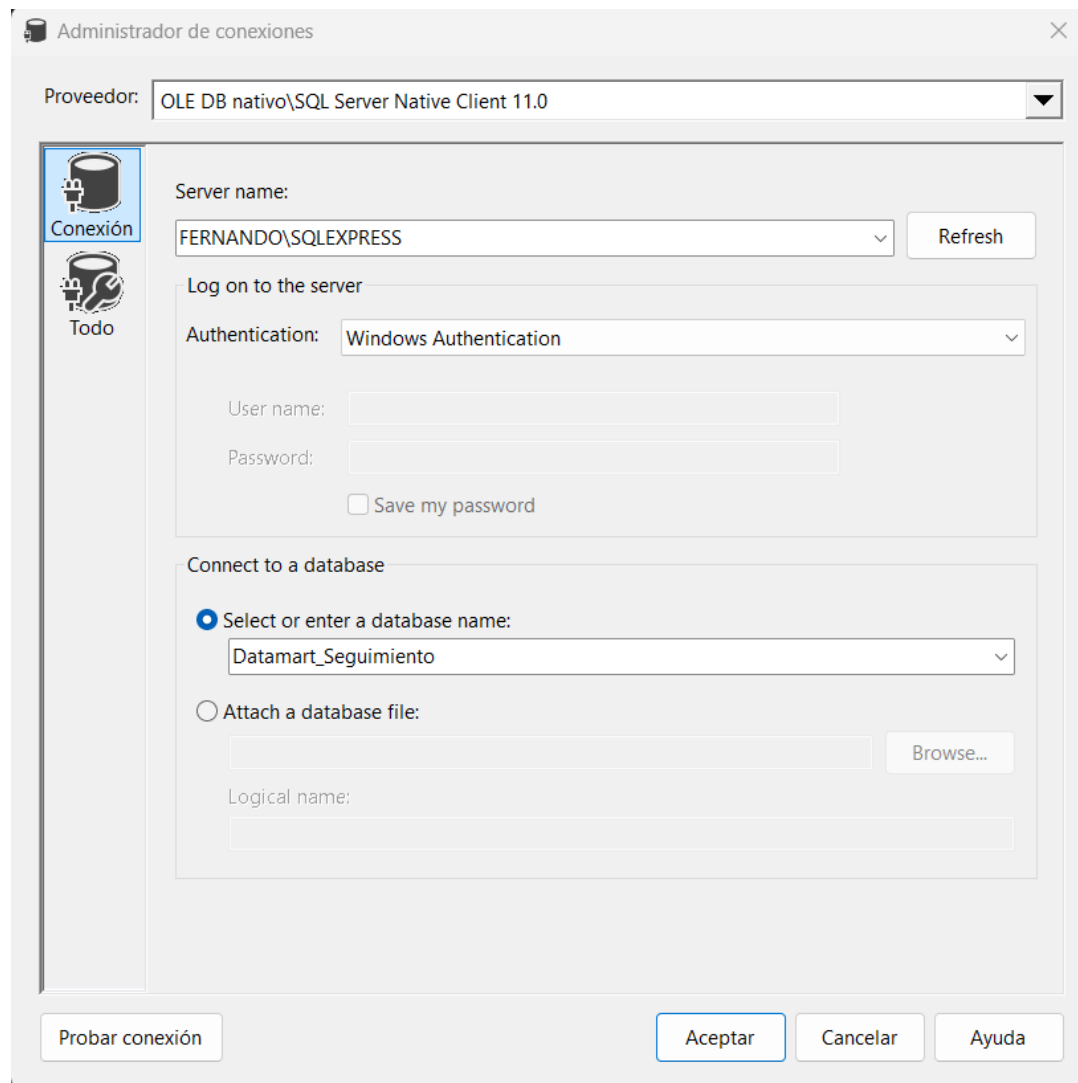
Para la implementación del cubo se utilizará el programa Visual Studio 2022, para el Analysis Service e implementación del mismo, para ejecutar el cubo.

a. Configuración de orígenes de datos

Se realizará las configuraciones para permitir la comunicación del SQL Server con el Analysis Service, así poder implementar el Datamart de seguimiento académico.

Figura 56

Administrador de la conexión con el Datamart



b. Construcción de la vista de datos

El asistente para cubos multidimensional nos solicitara las tablas que formaran parte de la vista, en la parte izquierda nos mostrara el listado de dimensiones y la tabla de hechos a seleccionar.

Figura 57

Selección de tablas para la generación de las vistas

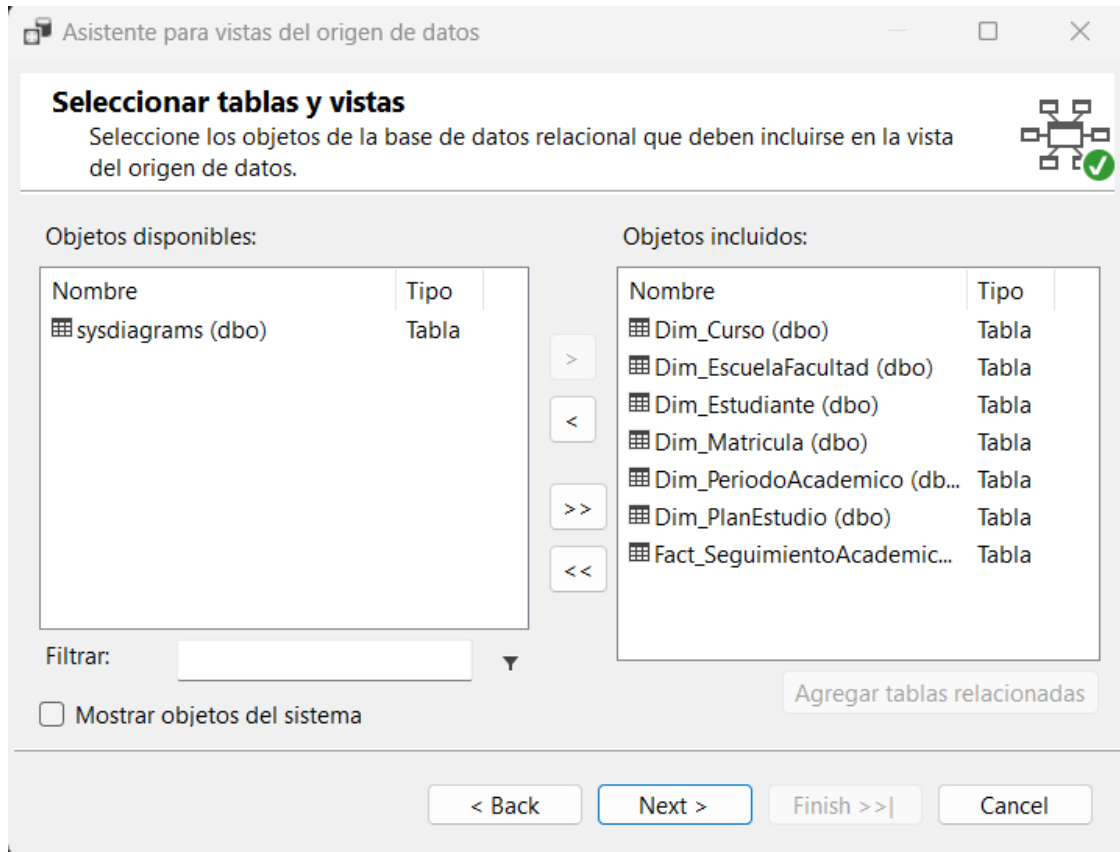


Figura 58

Dimensiones y tabla de hechos para generar la vista

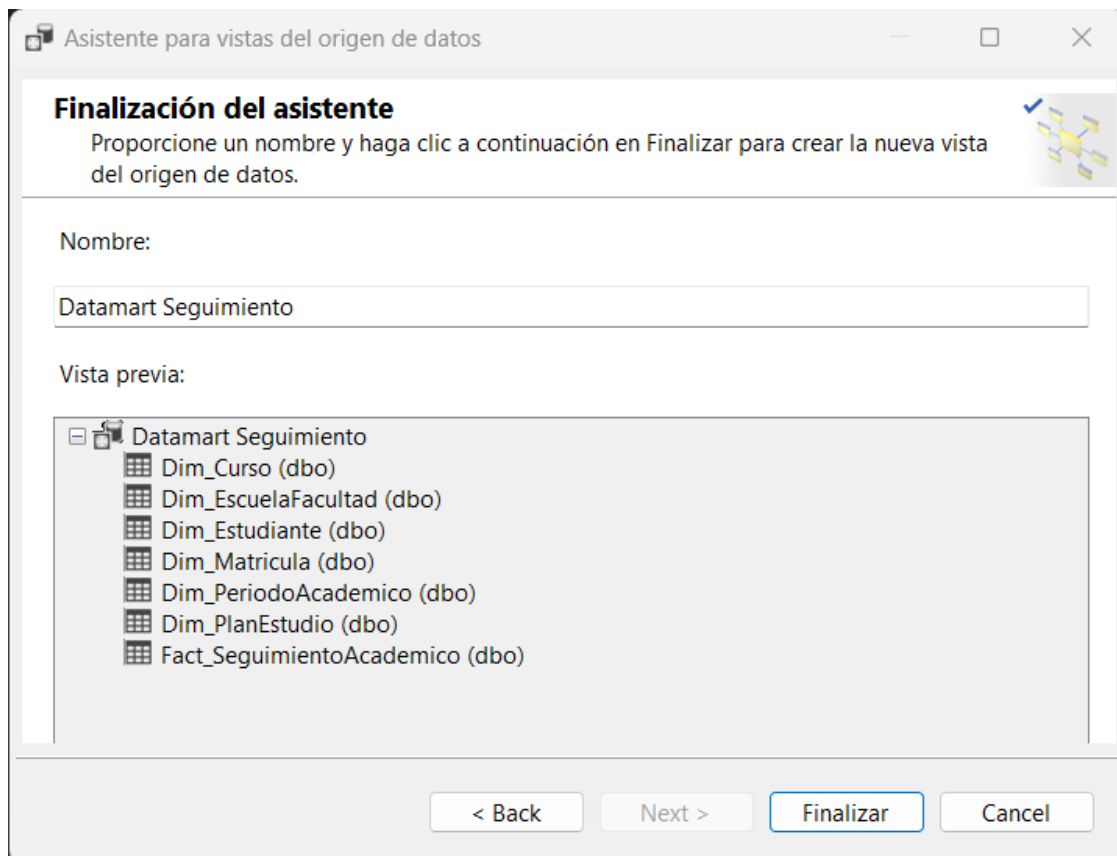
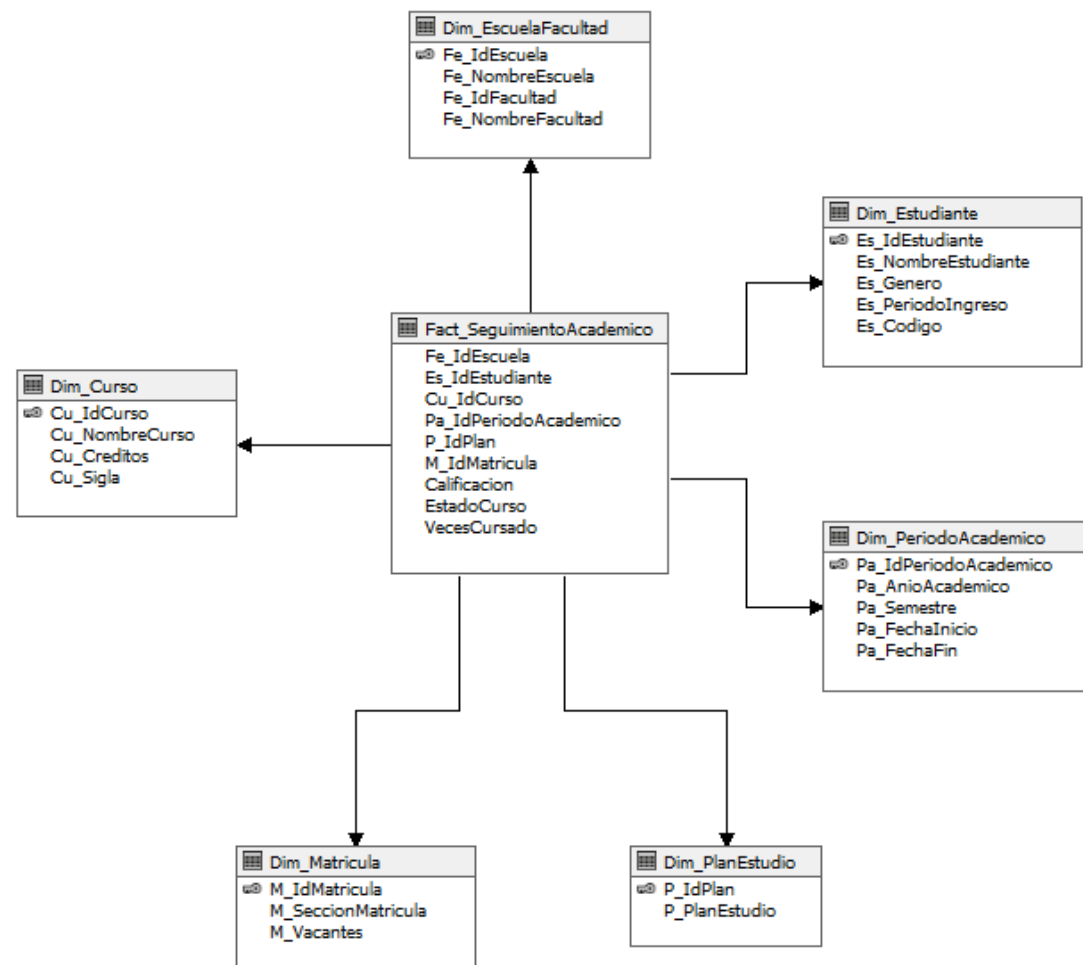


Figura 59

Vista de origen del esquema estrella del Datamart



La figura 59 muestra la vista creada, donde se visualiza las dimensiones y la tabla de hechos del Datamart, formando la primera vista; así como las uniones de las tablas de las dimensiones y la tabla de hechos en el esquema estrella.

c. Construcción del cubo multidimensional

Al tener las tablas de dimensiones y la tabla de hechos del Datamart, se realizará la carga en el asistente de cubos, para la generación de este.

Figura 60

Selección de la tabla principal

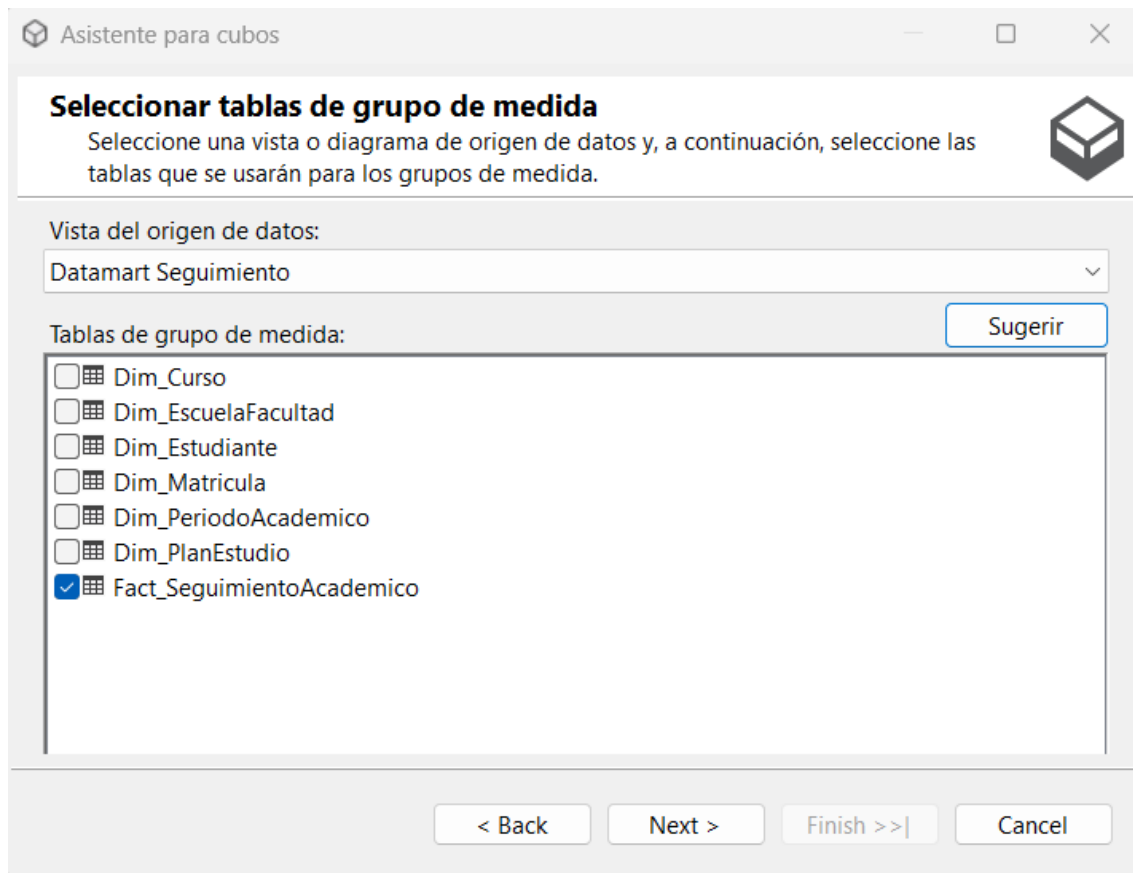


Figura 61

Tabla principal con sus medidas

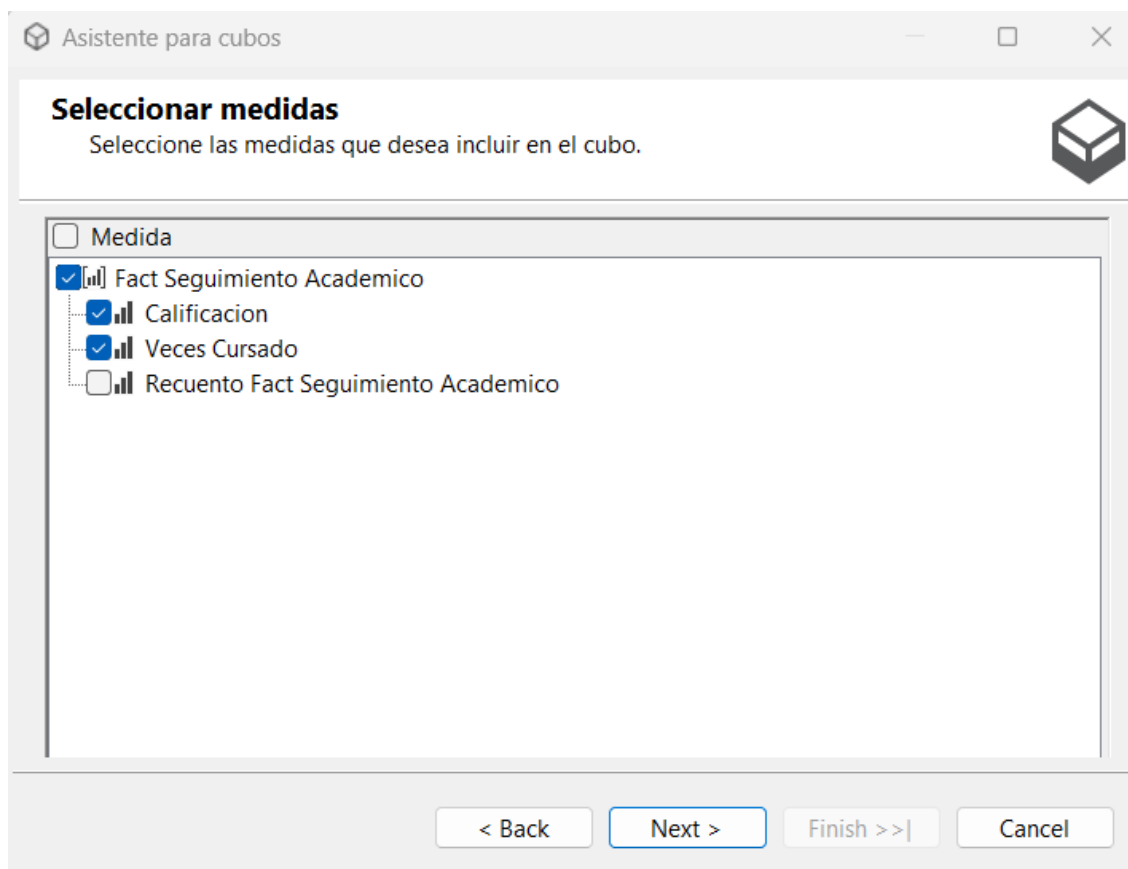


Figura 62

Creación del cubo dimensional y la relación con sus dimensiones

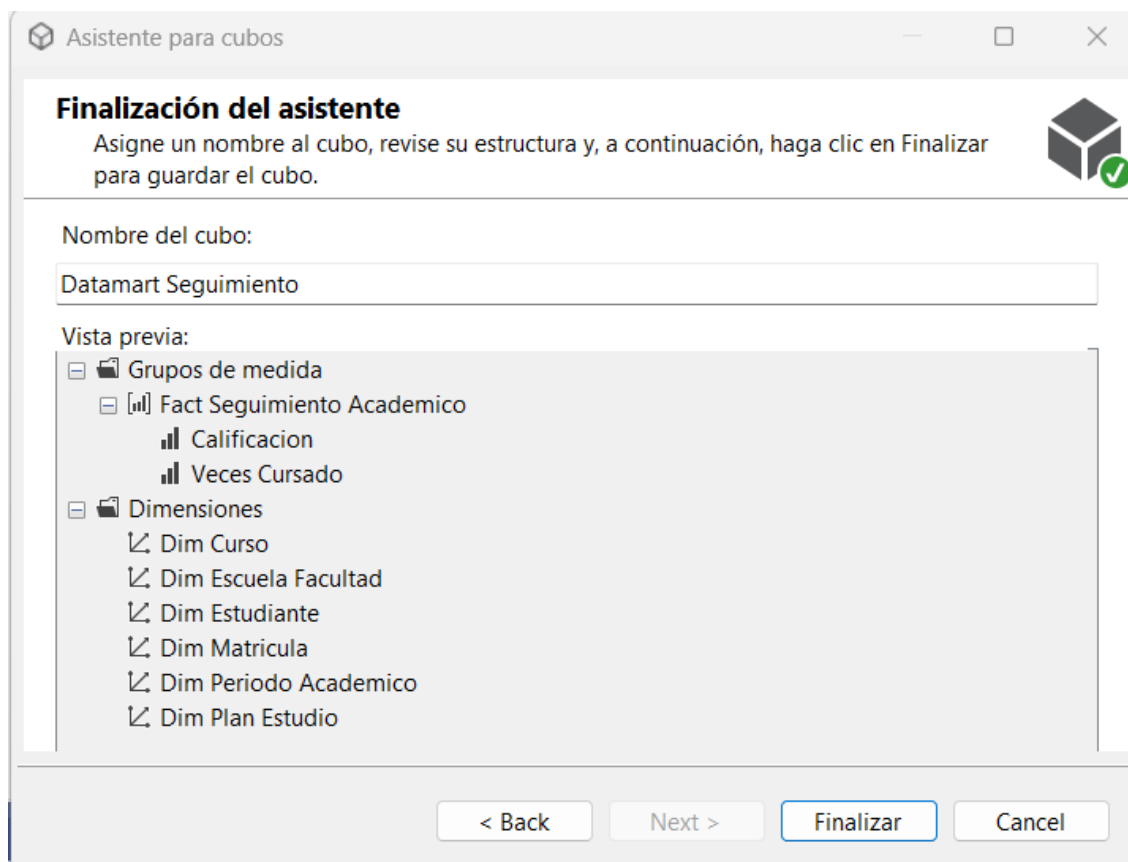
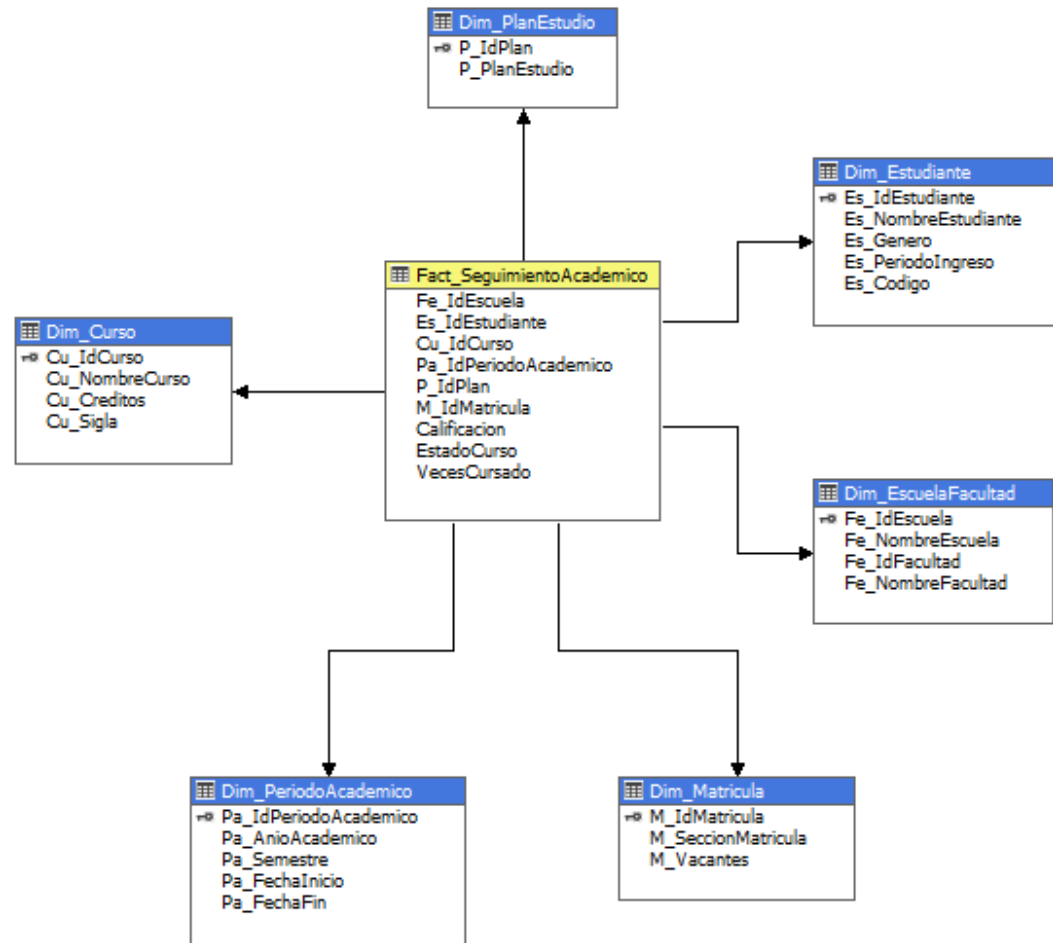


Figura 63

Diagrama final del cubo de Seguimiento académico

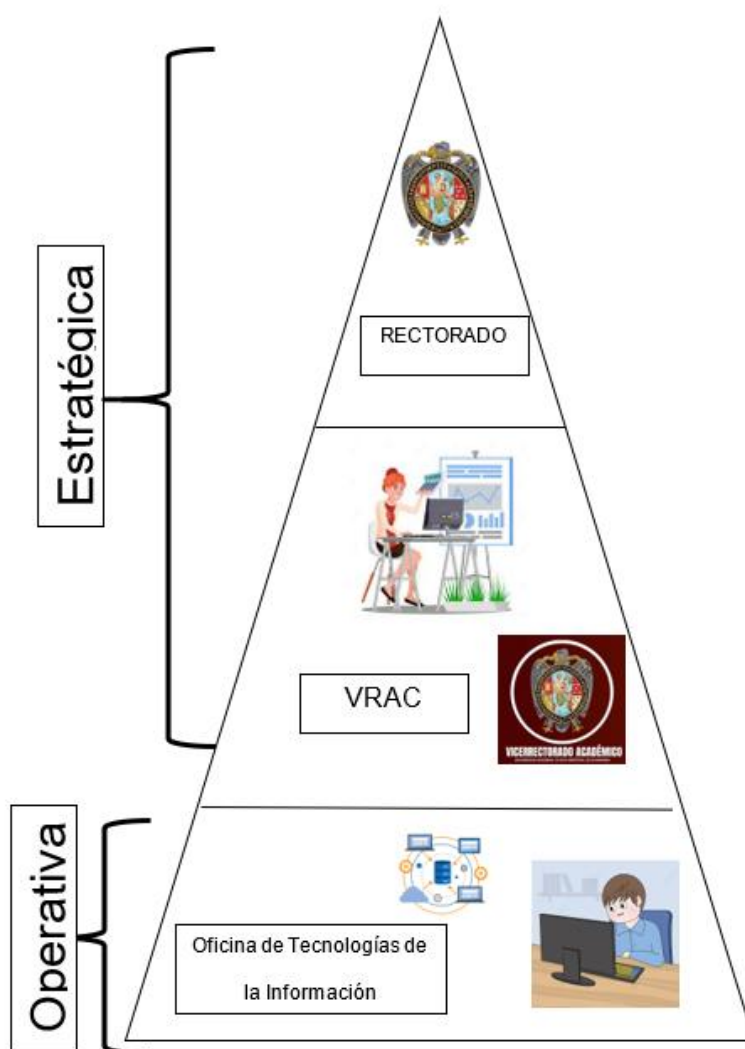


4.1.9. Especificación de aplicación para los usuarios finales

A continuación, se procede a desarrollar según la metodología Kimball, que usuarios involucrados en el modelo de negocios que serán los encargados de las tomas de decisiones, esto permitirá identificar los roles de la parte operativa y estratégica, tal como se muestra en la siguiente figura.

Figura 64

Usuarios asignados al Datamart de seguimiento académico



4.1.10. Desarrollo de la aplicación en el Power Bi

Con la creación del cubo multidimensional a continuación procederemos a crear las interfaces que ayudaran a los usuarios a entender la información, para lo cual usaremos la herramienta Power Bi para la creación de estas vistas.

a. Configuración para la conexión con el Power Bi

A continuación, se procederá a detallar la conexión necesaria de nuestro cubo multidimensional del Datamart de Seguimiento académico con la herramienta Power Bi para la creación de las vistas finales.

Figura 65

Configuración de la conexión con el Power Bi

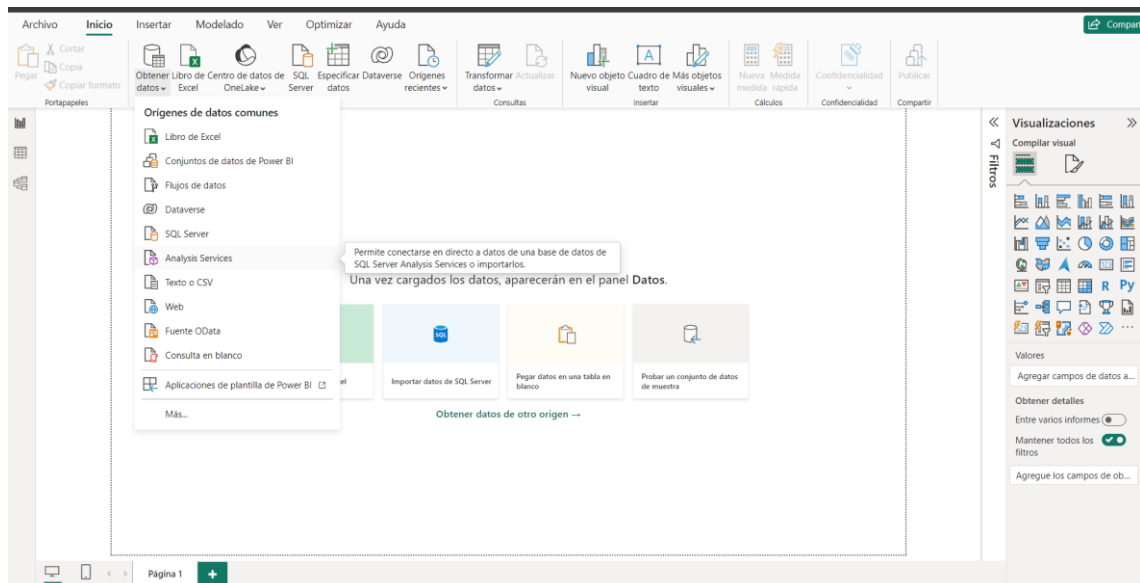
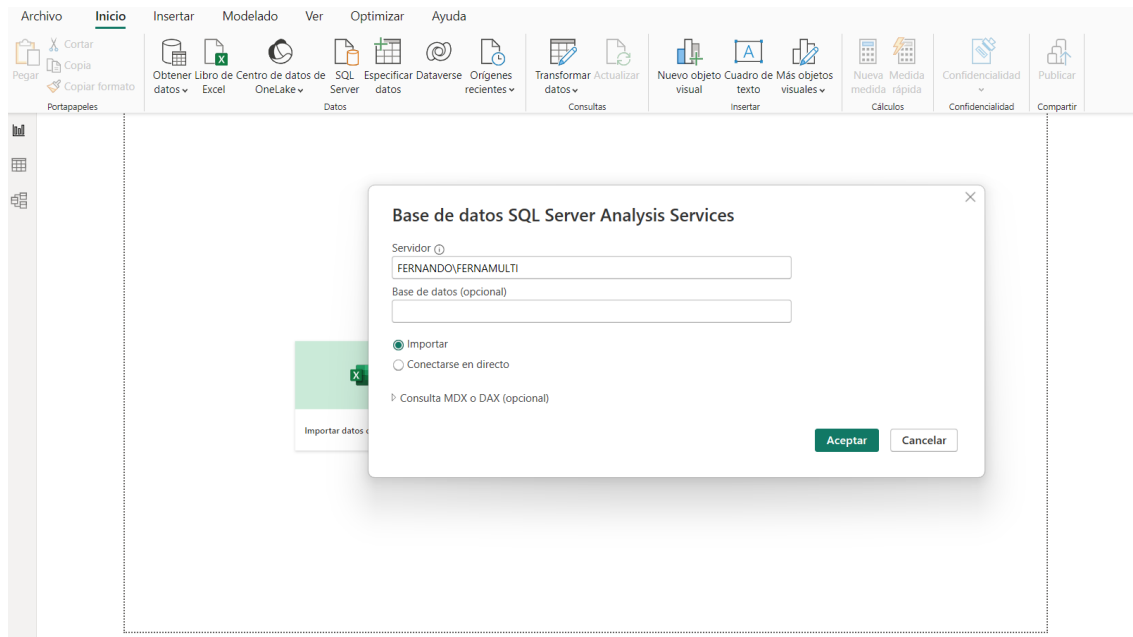


Figura 66

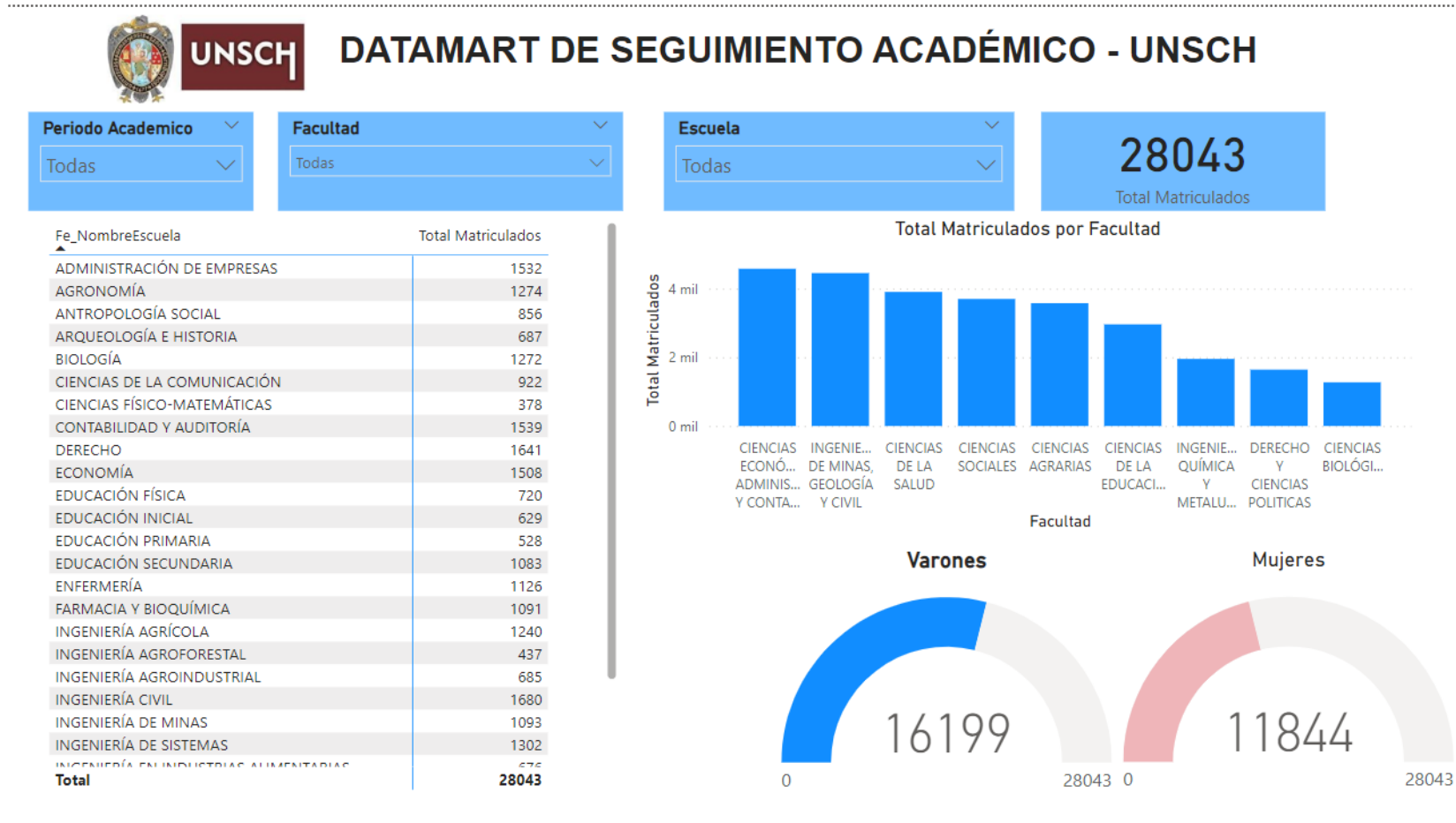
Configuración del servidor para la conexión con Power Bi



Vista del Power Bi, Datamart de Seguimiento Académico para el indicador cantidad de matriculados por género, escuela y facultad

Figura 67

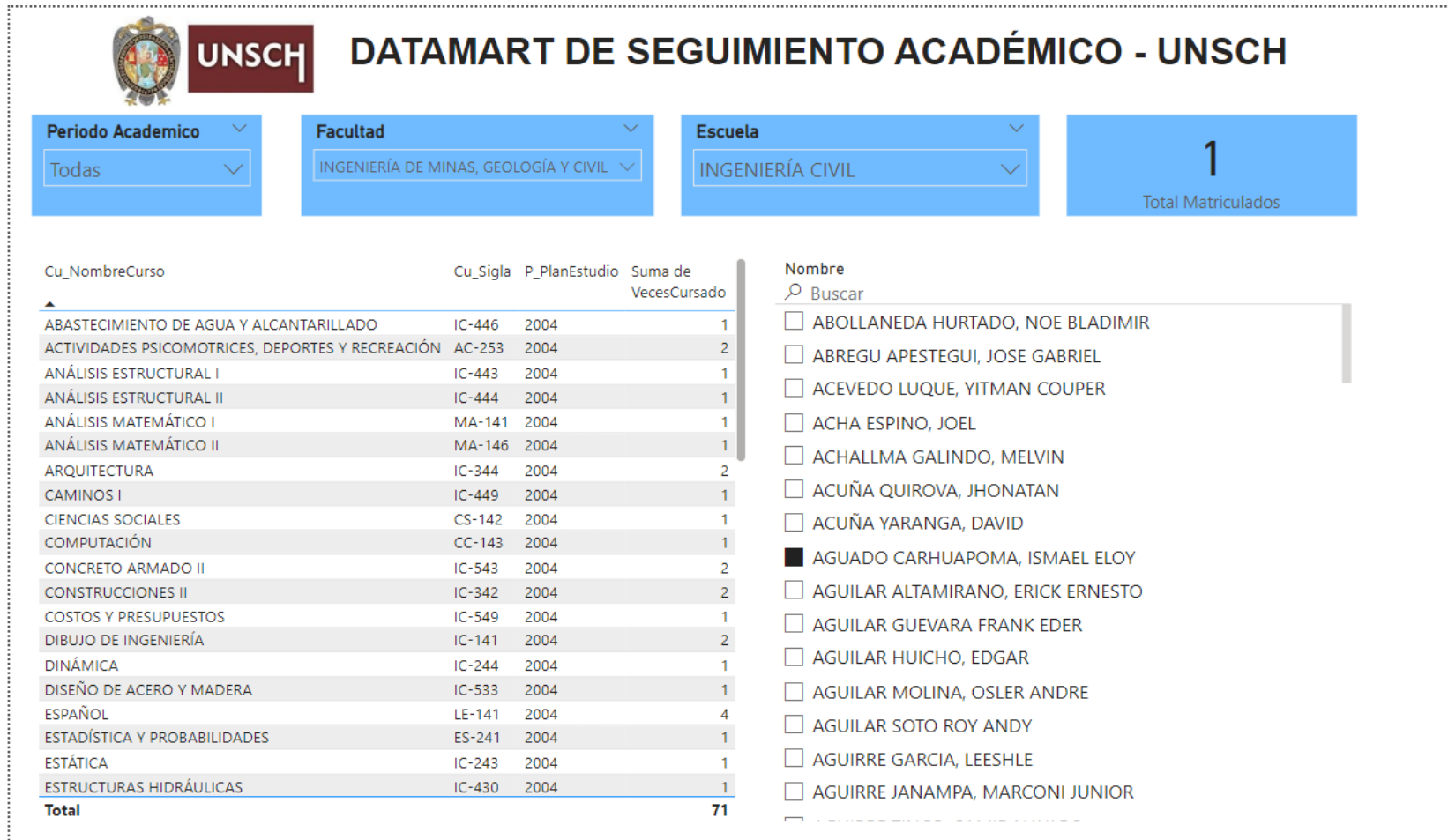
Cantidad de matriculados por género, escuela y facultad



Vista del Power Bi, Datamart de Seguimiento Académico para el indicador cantidad de veces que llevo un alumno sus cursos

Figura 68

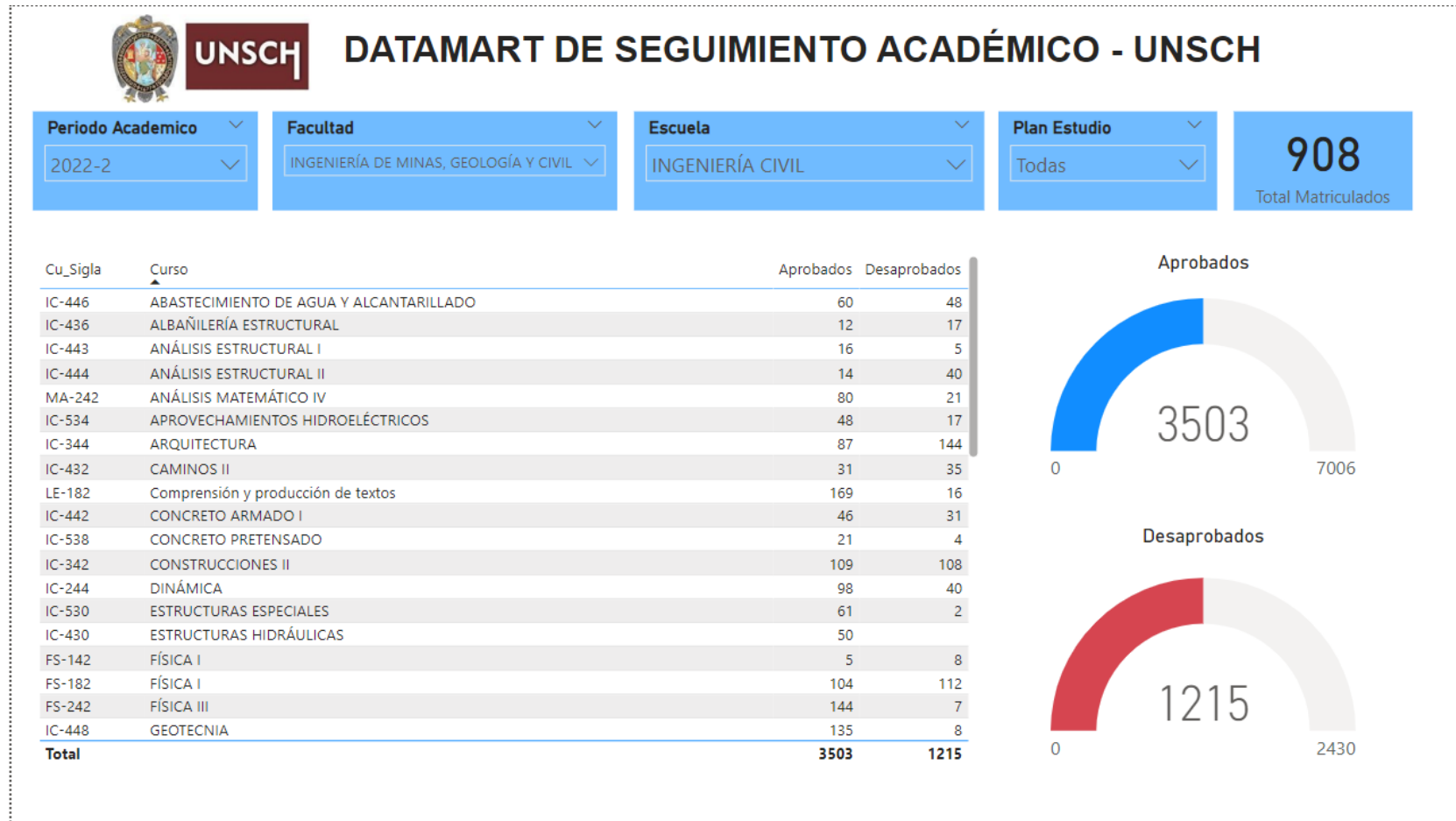
Cantidad de veces que llevo un alumno sus cursos



Vista del Power Bi, Datamart de Seguimiento Académico para el indicador cantidad de aprobados y desaprobados por curso, escuela y periodo

Figura 69

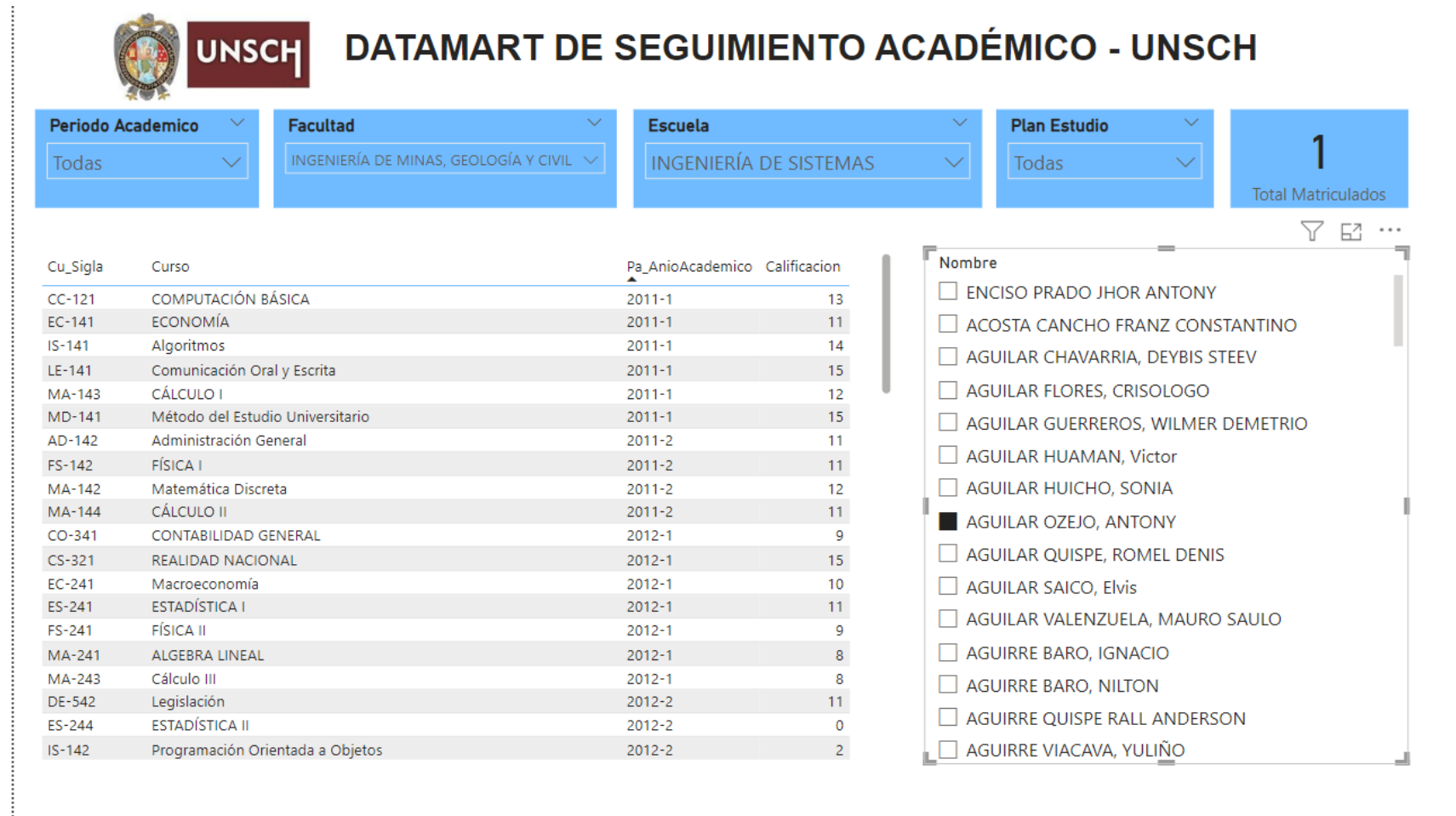
Cantidad de aprobados y desaprobados por curso, escuela y periodo



Vista del Power Bi, Datamart de Seguimiento Académico para el indicador cuales son las notas por curso y periodo académico del estudiante

Figura 70

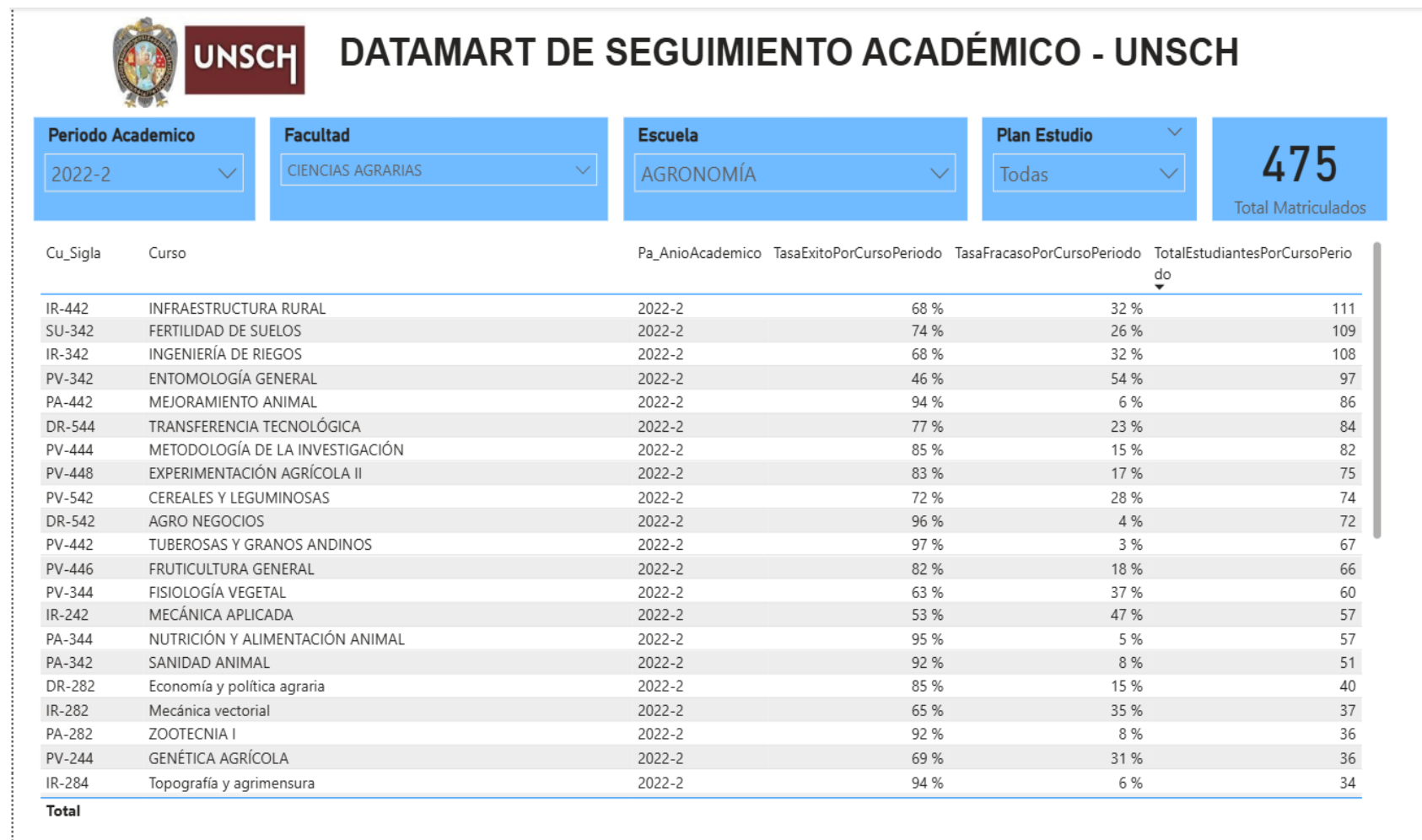
Notas por curso y periodo académico del estudiante



Vista del Power Bi, Datamart de Seguimiento Académico para el indicador tasa de éxito y fracaso de un curso por periodo

Figura 71

Tasa de éxito y fracaso de un curso por periodo



4.1.11. Crecimiento y mantenimiento del Datamart

a. Crecimiento del datamart

El desarrollo y expansión del Datamart son vitales para mejorar su operatividad y aprovechar al máximo el sistema de inteligencia de negocios. Aquí se muestra algunas propuestas para optimizar su rendimiento y aportar beneficios a la institución:

- **Incorporación de Datos Históricos y Futuros:** Añadir datos de años pasados y previsiones futuras permitirá realizar análisis más integrales y detectar tendencias a lo largo del tiempo.
- **Actualización del Nivel Analítico:** Introducir nuevos indicadores y datos de alto nivel ayudará a tomar decisiones estratégicas con información más precisa y relevante.
- **Creación de Nuevas Dimensiones:** Desarrollar nuevas dimensiones permitirá realizar análisis más detallados y específicos, adaptándose a las nuevas necesidades de la institución.
- **Establecimiento de Nuevas Tablas de Hechos:** Crear tablas de hechos adicionales, relacionadas con las existentes, permitirá generar nuevos informes conforme surjan nuevas necesidades. Esto aumentará la capacidad de respuesta ante demandas de información, beneficiando significativamente a la institución.

Con estas iniciativas, el Datamart no solo evolucionará en su capacidad de análisis, sino que también se convertirá en una herramienta más valiosa para la toma de decisiones informadas en la institución.

b. Mantenimiento del Datamart

Para mantener el Datamart actualizado se ha optado por actualizar semestralmente la base de datos ya que después del término de clases los docentes recién llenan las notas finales en su sistema, que es el insumo necesario para poder mostrar la información.

Capítulo V

Conclusiones y recomendaciones

5.1. Conclusiones

a. De acuerdo al marco teórico presentado en la presente investigación, se desarrolló el método para el diseño e implementación del DataMart de seguimiento académico para la universidad Nacional de San Cristóbal de Huamanga, se ha logrado los resultados que se muestra en las figuras 67 al 71, que brinda información académica significativa para la toma de decisiones, como son la tasa de éxito de los cursos, matrículas, notas por curso y reportes de rendimiento.

b. Usando la metodología Kimball, se mejoró la eficiencia del seguimiento con el Datamart desarrollado que era uno de los fines de la presente investigación, se ha logrado centralizar los datos y hacer la integración de los datos existentes, el cual se muestra en la figura 63.

c. En la presente investigación se diseñó el Datamart con el esquema estrella para garantizar la precisión de la información y la accesibilidad de los datos del seguimiento académico de los estudiantes, optimizando las consultas realizadas y facilitando el análisis de los datos tal como se muestra en la figura 32.

5.1. Recomendaciones

a. Se recomienda a los interesados en este trabajo de investigación profundizar en futuras investigaciones en el análisis predictivo del rendimiento estudiantil, donde se podrían implementar modelos estadísticos, que pronostiquen el éxito académico en función a más dimensiones de estudio.

b. Se recomienda realizar la inclusión de nuevas dimensiones para una visión más integral del seguimiento académico y profundizar en el rendimiento académico, para fomentar políticas que ayuden a los estudiantes con riesgo académico.

c. Se recomienda hacer capacitaciones periódicas a los usuarios del Datamart para que puedan aprovechar al máximo el uso de la herramienta y los reportes generados.

Referencias bibliográficas

- Cabanillas, K. G. (2011). *Análisis diseño e implementación de una solución de Inteligencia de Negocios para el área de compras y ventas de una empresa comercializadora de electrodomésticos*. [Tesis de pregrado]. Pontificia Universidad Católica del Perú, Lima, Perú.
- Bernabeu, R. D. (2010), *Metodología Hefesto* (5taEd.). Córdoba, Argentina: Editorial Tierra del sur.
- Carrasco, 2006. *Metodología de Investigación Científica*. Primera edición. Impreso en Perú.
- Castillo, G., Fernández, R., & Martínez, H. (2019). *Mejora de la calidad de los datos en sistemas de información académica mediante técnicas de limpieza y validación*. *Revista de Ingeniería de Software y Sistemas de Información*, 8(2), 35-49.
- Cobo, A. (s.f.). *Base de datos relacionales: Teoría y práctica* (1ª ed.). Madrid, España: Visión Libros.
- Durán, L. (2017) *Implementación de un Datamart para el seguimiento académico de los estudiantes en la escuela académico profesional de ingeniería de sistemas de la Universidad Nacional de Cajamarca*. [Tesis de grado, Universidad Nacional De Cajamarca]. Repositorio Institucional. <http://hdl.handle.net/20.500.14074/1144>
- Garatu Development. (s.f.). *Batch o lotes de producción*. Recuperado el 15 de octubre, 2022 de <https://development.grupogaratu.com/batch-lotes-produccion/>
- Gómez, A. y De Abajo, N. (1998). *Los sistemas de información en la empresa*. Madrid, España: Servicio de publicaciones de la Universidad de Oviedo.
- Hernández Sampieri, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación* (6ta Ed.). México, D.F., México: McGraw Hill Interamericana.
- Heurtel, O. (2009). *Php y MySQL Domine el Desarrollo de un sitio Web Dinámico e Interactivo* (1ra Ed.). Barcelona, España: Editorial ENI.
- Inmon, B. (2002), *Building the Data Warehouse* (3ra Ed.). Toronto. Canadá: Wiley.
- Joyanes, L. (2019). *Inteligencia de Negocios y analítica de datos*. México: Alfa omega.
- JVS Informática Blog. (s. f.). *Procesamiento por lotes o batch*. Recuperado el 15 de octubre, 2022 de <https://www.jvs-informatica.com/blog/glosario/procesamiento-por-lotes-batch/>
- Kimball, R. y Ross, M. (2013). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (2da edición). Toronto. Canadá: Wiley.

- Laudon, K. y Laudon, J. (2008). *Sistemas de Información Gerencial* (10ma Ed.). México, D.F., México: Pearson Educación.
- Loshin, D. (2012). *The Savvy Manager's Guide - MK Series on Business Intelligence*. Newnes
- Luque, I., Gómez, M., López, N. y Cerruela, G. (2002). *Base de Datos*. México D.F., México: Alfaomega Grupo Editor.
- Moss, L. y Atre, S. (2003) *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Boston. Inglaterra: Addison Wesley.
- Nader, J. (2003). *Sistema de Apoyo gerencial Universitario*. [Tesis de postgrado]. Instituto Tecnológico de Buenos Aires. Buenos Aires. Argentina.
- Nima, J. (2009). *Soluciones OLAP con Microsoft SQL Server AnalysisServices*. Recuperado de <http://www.eumed.net/libros-gratis/2009c/574/index.htm>.
- Oppel, A. y Sheldon, R. (2009). *Fundamentos de SQL* (3ª Ed.). México D.F., México: McGraw Hill Interamericana.
- Osorio, F. (2008). *Base de datos relacionales: Teoría y práctica* (1ª ed.). Madrid, España: Thomson.
- Pérez, J. del C., Samaniego, F.F. (2014). *Sistema de evaluación y seguimiento del rendimiento académico*. Facultad de Ingeniería de Sistemas Computacionales. Centro Regional de Azuero. Universidad Tecnológica de Panamá. Prisma Tecnológico. http://www.utp.ac.pa/documentos/2015/pdf/04-SISTEMA_EVALUACION_16-19_0.pdf
- Puryear, B. (1993). *Data Warehousing Fundamentals*. Prentice Hall.
- Ramos, S. (2016). *Data Warehouse, Data Marts y Modelos Dimensionales. Un pilar fundamental para la toma de decisiones*. Alicante: Solid Q. Press.
- Reyes, L. (2004). *El bajo rendimiento académico de los estudiantes universitarios. Una Aproximación a sus causas*. Revista Theorethiko, Año VI, Recuperado de, <http://www.ufg.edu.sv/ufg/theorethikos/Junio04/ebr.html>
- Rivadera, G. (s.f) *Metodología Kimball para el diseño de almacenes de datos (Data Warehouse)* Lima, Perú.
- Rodríguez, M. (2018) *Análisis y diseño de un Datamart para el seguimiento académico de alumnos en un entorno universitario*. [Tesis de grado, Universidad Carlos III de Madrid]. Repositorio Institucional. <http://hdl.handle.net/10016/9856>

- Ruiz, E., Ruiz, G., Odstrcil, M. (2007). *Metodología para realizar el seguimiento académico de alumnos universitarios*. Facultad de Ciencias Exactas y Tecnología (FACET), Universidad Nacional de Tucumán (UNT), Argentina. OEI - Revista Iberoamericana de Educación. Recuperado de, <http://www.rieoei.org/deloslectores/1590Ruiz.pdf>
- Sarmiento, H. (2018) *Inteligencia de negocios usando Pentaho para la gestión académica en la UNAMBA-2016*. [Tesis de grado, Universidad Nacional Micaela Bastidas de Apurímac]. Repositorio académico UNAMBA-Institucional. <http://repositorio.unamba.edu.pe/handle/UNAMBA/613>
- Silberschatz, A., Korth H. y Sudarshan S. (2005). *Fundamentos de bases de datos* (4ta Ed.). España, Madrid: McGraw Hill Interamericana.
- Sistrix. (s.f.). *¿Qué es una búsqueda transaccional?* Recuperado el 15 de octubre, 2022 de <https://www.sistrix.es/preguntale-a-sistrix/que-es-una-busqueda-transaccional/>
- Supo, J. (2015). *Taxonomía de la investigación: El arte de clasificar aplicado a la investigación científica*.
- Tamayo, M. y Moreno, F. (2006, diciembre) *Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP*. Ingeniería e Investigación. Recuperado el 16 de febrero del 2018, de <http://www.redalyc.org/html/643/64326317/>
- Textbroker. (s.f.). *Tipos de búsqueda: transaccionales, de navegación, informativas*. Recuperado el 15 de octubre, 2022 de <https://www.textbroker.es/tipos-de-busqueda-transaccionales-de-navegacion-informativas#:~:text=Las%20b%C3%BAsquedas%20transaccionales%20son%20consultas, encontrar%20informaci%C3%B3n%20relacionada%20con%20transacciones.>
- Torres, L. (2007). *Business Intelligence*. Recuperado de <http://www.gravitar.biz/index.php/bi/bi-terminologia-1>.
- Ullman, J.D. y Widom, J. (1999). *Introducción a los Sistemas de Bases de Datos*. México D.F., México: Editorial PHH.
- Valdiviezo, M. J., Herrera, I. Z. S. y Jáuregui, G. D. (2007). *Análisis y Diseño de una Herramienta de Desarrollo de Soluciones para Inteligencia de Negocios – Análisis dimensional*. [Tesis de pregrado]. Pontificia Universidad Católica del Perú. Lima. Perú.

- Vitt, E. y Luckevich, M. (2002). *Business Intelligence Técnicas de análisis para la toma de decisiones estratégicas* (1era Ed.). Madrid. España: McGraw-Hill/Interamericana.
- Yalán, J. y Palomino, L. (2013, enero - junio), *Implementación de un Datamart como una solución de Inteligencia de Negocios para el área de logística de T-Impulso*, Revista de investigación de sistemas e informática, Universidad Nacional Mayor de San Marcos, Lima, Perú.
- Zambrano, J. (2011). *Análisis, diseño e implementación de un datamart para el área de mantenimiento y logística de una empresa de transporte público de pasajeros*. [Tesis de pregrado]. Pontificia Universidad Católica del Perú, Lima, Perú.

Anexo 1: Operacionalización de variables

Variable	Dimensión	Indicador	Ítem
Datamart	Diseño y Arquitectura	Estructura técnica del Datamart	¿El Datamart tiene una estructura que permite la consulta eficiente de la base de datos?
			¿El diseño del Datamart incluye mecanismos para mantener la integridad de los datos?
	Integración de Datos	Datos integrados	¿El Datamart integra datos de distintas fuentes académicas (registros de matrículas, calificaciones, etc.)? ¿El Datamart permite la integración de datos en tiempo real, cuando se implemente en línea?
	Calidad de los Datos	Precisión y confiabilidad de los datos	¿El Datamart permite asegurar la precisión de los datos?
Seguimiento Académico	Eficiencia del Seguimiento	Rapidez del acceso a la información	¿El Datamart permite generar reportes de seguimiento académico de manera rápida?

		¿El sistema ofrece consultas en tiempo real sobre el rendimiento académico?
Precisión de la Información	Fiabilidad de los datos académicos	¿Los datos académicos proporcionados por el Datamart son precisos?
Accesibilidad de los Datos		¿El Datamart facilita el acceso a la información académica por parte de los usuarios autorizados?

Anexo 2: Fichas de análisis para la BD de la universidad

Ficha 1:

¿Cuántos matriculados hay en un periodo por género, escuela y facultad?

```
SELECT DISTINCT
    efa.Name AS Facultad,
    gca.Name AS Escuela,
    ete.Name AS Periodo,
    COUNT(DISTINCT iah.StudentId) AS CantidadMatriculados
FROM [db_sigau].[unsch.sigau.db].[intranet_academichistories] iah
INNER JOIN [unsch.sigau.db].[generals_students] gstu ON gstu.Id = iah.StudentId
INNER JOIN [unsch.sigau.db].[enrollment_terms] ete ON ete.Id = iah.TermId
INNER JOIN [unsch.sigau.db].[generals_careers] gca ON gca.Id = gstu.CareerId
INNER JOIN [unsch.sigau.db].[enrollment_faculties] efa ON efa.Id = gca.FacultyId
WHERE iah.Type=1
GROUP BY gca.Name, efa.Name, ete.Name;
```

Ficha 2:

¿Cantidad de veces llevo un alumno sus cursos?

```
SELECT DISTINCT
    efa.Name AS Facultad,
    gca.Name AS Escuela,
    ete.Name AS Periodo,
    aus.FullName as Nombre,
    ecu.Name as Curso,
    iah.Try AS VecesCurso
FROM [db_sigau].[unsch.sigau.db].[intranet_academichistories] iah
INNER JOIN [unsch.sigau.db].[generals_students] gstu ON gstu.Id = iah.StudentId
INNER JOIN [unsch.sigau.db].aspnetusers aus on aus.Id=gstu.UserId
INNER JOIN [unsch.sigau.db].[enrollment_terms] ete ON ete.Id = iah.TermId
INNER JOIN [unsch.sigau.db].[generals_careers] gca ON gca.Id = gstu.CareerId
INNER JOIN [unsch.sigau.db].[enrollment_faculties] efa ON efa.Id = gca.FacultyId
INNER JOIN [unsch.sigau.db].enrollment_courses ecu on ecu.Id=iah.CourseId
WHERE ete.Name = '2019-1' and iah.Type=1
GROUP BY gca.Name, efa.Name, ete.Name, aus.FullName, iah.Try, ecu.Name;
```

Ficha 3:

¿Cantidad de aprobados y desaprobados por curso y escuela durante un periodo?

```
SELECT
    efa.Name AS Facultad,
    gca.Name AS Escuela,
    ete.Name AS Periodo,
    ecu.Name AS Curso,
    SUM(CASE WHEN iah.Approved = 1 THEN 1 ELSE 0 END) AS CantidadAprobados,
    SUM(CASE WHEN iah.Approved = 0 THEN 1 ELSE 0 END) AS CantidadDesaprobados
FROM [db_sigau].[unsch.sigau.db].[intranet_academichistories] iah
INNER JOIN [unsch.sigau.db].[generals_students] gstu ON gstu.Id = iah.StudentId
INNER JOIN [unsch.sigau.db].[aspnetusers] aus ON aus.Id = gstu.UserId
INNER JOIN [unsch.sigau.db].[enrollment_terms] ete ON ete.Id = iah.TermId
INNER JOIN [unsch.sigau.db].[generals_careers] gca ON gca.Id = gstu.CareerId
INNER JOIN [unsch.sigau.db].[enrollment_faculties] efa ON efa.Id = gca.FacultyId
INNER JOIN [unsch.sigau.db].[enrollment_courses] ecu ON ecu.Id = iah.CourseId
WHERE ete.Name = '2019-1' AND iah.Type = 1
GROUP BY gca.Name, efa.Name, ete.Name, ecu.Name;
```

Ficha 4:

¿Cuáles son las notas por curso y periodos académicos cursados por el estudiante?

```
SELECT
    efa.Name AS Facultad,
    gca.Name AS Escuela,
    ete.Name AS Periodo,
    aus.FullName as Nombres,
    ecu.Name AS Curso,
    iah.Grade
FROM [db_sigau].[unsch.sigau.db].[intranet_academichistories] iah
INNER JOIN [unsch.sigau.db].[generals_students] gstu ON gstu.Id = iah.StudentId
INNER JOIN [unsch.sigau.db].[aspnetusers] aus ON aus.Id = gstu.UserId
INNER JOIN [unsch.sigau.db].[enrollment_terms] ete ON ete.Id = iah.TermId
INNER JOIN [unsch.sigau.db].[generals_careers] gca ON gca.Id = gstu.CareerId
INNER JOIN [unsch.sigau.db].[enrollment_faculties] efa ON efa.Id = gca.FacultyId
INNER JOIN [unsch.sigau.db].[enrollment_courses] ecu ON ecu.Id = iah.CourseId
WHERE ete.Name = '2019-1' AND iah.Type = 1
GROUP BY gca.Name, efa.Name, ete.Name, ecu.Name, iah.Grade, aus.FullName;
```

Anexo 3: Guía de entrevista para el Datamart

“Datamart para seguimiento académico de los estudiantes en la Universidad Nacional De San Cristóbal De Huamanga, Ayacucho, 2022.”

Fecha:

Entrevistado:

Preguntas:

¿Cuál es la capacidad del Datamart para realizar consultas rápidas y eficientes sobre sus datos?

1. ¿El diseño del Datamart garantiza la corrección y consistencia de los datos durante las actualizaciones?

2. ¿El Datamart puede combinar datos de distintas fuentes como registros de matrículas y calificaciones? ¿Cómo se realiza esa integración?

3. Una vez en línea, ¿crees que el Datamart podrá procesar e integrar datos en tiempo real? ¿Qué ventajas ofrecería esta funcionalidad?

4. ¿Consideras que el Datamart asegura la exactitud de los datos? ¿Qué mecanismos existen para garantizar esta precisión?

5. Desde tu experiencia, ¿el Datamart facilita la creación de informes académicos de manera eficiente? ¿Cuánto tiempo suele tomar generar un informe típico?

6. ¿El sistema permite consultas en tiempo real sobre el rendimiento académico de los estudiantes? ¿Cómo influye esto en la toma de decisiones?

7. En términos generales, ¿crees que los datos proporcionados por el Datamart son fiables y precisos? ¿Has detectado algún error o inconsistencia en los datos?

8. ¿Cómo es el proceso de acceso a la información académica en el Datamart? ¿Crees que el sistema facilita el acceso a los usuarios autorizados?



UNSCH

FACULTAD DE
INGENIERÍA
DE MINAS, GEOLOGÍA Y CIVIL

ACTA DE SUSTENTACIÓN DE TESIS N° 01-2026-FIMGC

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

En la Universidad Nacional de San Cristóbal de Huamanga, en la ciudad de Ayacucho, en cumplimiento a la **Resolución Decanal No 021-2026-FIMGC-D**, a los **diecisiete días del mes de marzo de 2026**, siendo las **10:00 a.m.**, reunidos en el **Auditorio de la Escuela Profesional de Ingeniería de Minas**, bajo la presidencia del **MSc. Ing. José Ernesto ESTRADA CÁRDENAS**, y los miembros: **Mg. Ing. Karel PERALTA SOTOMAYOR**, **Dr. Ing. Hubner JANAMPA PATILLA** y **Mg. Ing. Javier PORTILLO QUISPE**, actuando como secretario docente el **MSc. Ing. Saul Walter RETAMOZO FERNANDEZ**, para proceder a la sustentación de tesis para optar el **Título Profesional de Ingeniero de Sistemas**, del bachiller en Ingeniería de Sistemas:

Fernando ANAYA POMA

Quien presentó la tesis denominada:

Datamart para el seguimiento académico de los estudiantes en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024

Los señores miembros del jurado luego de expuesta la tesis y absueltas las preguntas, deliberaron y declararon:

Aprobado con 16 (diesiséis)

Siendo las **11:36 a.m.** del día **17 de marzo del 2026**, culmina el acto de sustentación de tesis, y en conformidad de lo actuado los miembros del jurado firmamos al pie del presente.

MSc. Ing. José Ernesto ESTRADA CÁRDENAS
Presidente

Mg. Ing. Karel PERALTA SOTOMAYOR
Miembro

Mg. Ing. Javier PORTILLO QUISPE
Miembro

Dr. Ing. Hubner JANAMPA PATILLA
Miembro - Asesor

MSc. Ing. Saul Walter RETAMOZO FERNANDEZ
Secretario docente de la FIMGC

FACULTAD DE INGENIERÍA
DE MINAS Y CIVIL
Av. Independencia S/N
Ciudad Universitaria
Central Tel. 066 312510
Anexo 151



UNSCH

FACULTAD DE
INGENIERÍA
DE MINAS, GEOLOGÍA Y CIVIL



CONSTANCIA DE ORIGINALIDAD DE TRABAJO DE INVESTIGACIÓN

CONSTANCIA N° 034-2026-KPS-FIMGC/UNSCH

El que suscribe; responsable verificador de originalidad de trabajos de tesis de pregrado con el software Turnitin, en segunda instancia para las **Escuelas Profesionales** de la **Facultad de Ingeniería de Minas, Geología y Civil**; en cumplimiento a la **Resolución de Consejo Universitario N° 039-2021-UNSCH-CU**, Reglamento de Originalidad de Trabajos de Investigación de la Universidad Nacional San Cristóbal de Huamanga y **Resolución Decanal N° 697-2024-FIMGC-D**, deja constancia de originalidad de trabajo de investigación, que el/la Sr./Srta.

Nombres y Apellidos : Fernando Anaya Poma
Escuela Profesional : INGENIERÍA DE SISTEMAS
Título de la Tesis : Datamart para el seguimiento académico de los estudiantes en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024
Evaluación de la Originalidad : **21%** Índice de Similitud
Identificador de la entrega : 2927225729

Por tanto, según los Artículos 12, 13 y 17 del Reglamento de Originalidad de Trabajos de Investigación, es **PROCEDENTE** otorgar la **Constancia de Originalidad** para los fines que crea conveniente.

En señal de conformidad y verificación se firma la presente constancia

Ayacucho, 10 de abril de 2026



Firmado digitalmente por:
PERALTA SOTOMAYOR Karel
FAU 20143680754 soft
Motivo: Soy el autor del documento
Fecha: 10/04/2026 13:51:39-0500

Datamart para el seguimiento académico de los estudiantes en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024

por Fernando Anaya Poma

Fecha de entrega: 09-abr-2026 08:13p. m. (UTC-0500)

Identificador de la entrega: 2927225729

Nombre del archivo: MEMORANDO_Nº_205-2026-FIMGC-UNSCH-CERTIFICADO_DE_ORIGINALIDAD-FERNANDO_ANAYA_POMA.pdf (5.94M)

Total de palabras: 15070

Total de caracteres: 83526

Datamart para el seguimiento académico de los estudiantes en la Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, 2024

INFORME DE ORIGINALIDAD

21%

INDICE DE SIMILITUD

22%

FUENTES DE INTERNET

5%

PUBLICACIONES

13%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	Submitted to Universidad Nacional de San Cristóbal de Huamanga Trabajo del estudiante	8%
2	repositorio.unsch.edu.pe Fuente de Internet	8%
3	hdl.handle.net Fuente de Internet	2%
4	repositorio.autonoma.edu.pe Fuente de Internet	1%
5	qdoc.tips Fuente de Internet	1%
6	1library.co Fuente de Internet	<1%
7	Odicio Ortiz, Walter Miguel. "Solución holística basada en inteligencia de negocios como herramienta de soporte para optimizar la toma de decisiones a nivel estratégico en el	<1%

gobierno regional Piura en el año 2016",
Universidad Católica los Ángeles de Chimbote
(Peru)

Publicación

8	apirepositorio.unh.edu.pe Fuente de Internet	<1 %
9	docplayer.es Fuente de Internet	<1 %
10	repositorio.upao.edu.pe Fuente de Internet	<1 %
11	repositorio.autonoma.edu.co Fuente de Internet	<1 %
12	core.ac.uk Fuente de Internet	<1 %

Excluir citas

Activo

Excluir coincidencias

< 30 words

Excluir bibliografía

Activo