



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

**SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO  
ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS  
DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE  
GUATEMALA**

**Sandra Eunice Jimenez Rodas**

Asesorado por la Inga. Romelia Angelina Jiménez Rodas

Guatemala, julio de 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO  
ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS  
DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE  
GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**SANDRA EUNICE JIMENEZ RODAS**

ASESORADO POR LA INGA. ROMELIA ANGELINA JIMÉNEZ RODAS

AL CONFERÍRSELE EL TÍTULO DE

**INGENIERA EN CIENCIAS Y SISTEMAS**

GUATEMALA, JULIO DE 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANO	Ing. José Francisco Gómez Rivera (a.i.)
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton De León Bran
VOCAL IV	Ing. Kevin Vladimir Armando Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Inga. Floriza Felipa Ávila Pesquera de Medinilla
EXAMINADOR	Ing. Sergio Leonel Gómez Bravo
EXAMINADOR	Ing. Carlos Alfredo Azurdia Morales
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

## **HONORABLE TRIBUNAL EXAMINADOR**

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha 05 de febrero de 2022.



**Sandra Eunice Jimenez Rodas**

Universidad de San Carlos de  
Guatemala



Facultad de Ingeniería  
Unidad de EPS

Guatemala, 24 de febrero de 2023.  
REF.EPS.DOC.104.02.2023.

Ing. Oscar Argueta Hernández  
Director Unidad de EPS  
Facultad de Ingeniería  
Presente

Estimado Ingeniero Argueta Hernández:

Por este medio atentamente le informo que como Supervisora de la Práctica del Ejercicio Profesional Supervisado, (E.P.S) de la estudiante universitaria de la Carrera de Ingeniería en Ciencias y Sistemas, **Sandra Eunice Jimenez Rodas, Registro Académico 201408470 y CUI 2976 03787 0101** procedí a revisar el informe final, cuyo título es **SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA.**

En tal virtud, **LO DOY POR APROBADO**, solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,

“Id y Enseñad a Todos”



Inga. Floriza Felipa Ávila Pesquera de Medinilla  
Supervisora de EPS  
Área de Ingeniería en Ciencias y Sistemas

FFAPdM/RA

Universidad de San Carlos de  
Guatemala



Facultad de Ingeniería  
Unidad de EPS

Guatemala, 24 de febrero de 2023.  
REF.EPS.D.69.02.2023.

Ing. Carlos Gustavo Alonzo  
Director Escuela de Ingeniería Ciencias y Sistemas  
Facultad de Ingeniería  
Presente

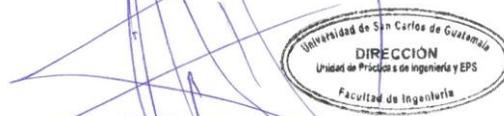
Estimado Ingeniero Alonzo:

Por este medio atentamente le envío el informe final correspondiente a la práctica del Ejercicio Profesional Supervisado, (E.P.S) titulado **SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, que fue desarrollado por la estudiante universitaria **Sandra Eunice Jimenez Rodas, Registro Académico 201408470 y CUI 2976 03787 0101** quien fue debidamente asesorada por la Inga. Romelia Angelina Jiménez Rodas y supervisada por la Inga. Floriza Felipa Ávila Pesquera de Medinilla.

Por lo que habiendo cumplido con los objetivos y requisitos de ley del referido trabajo y existiendo la aprobación del mismo por parte de la Asesora y la Supervisora de EPS, en mi calidad de Director apruebo su contenido solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,  
"Id y Enseñad a Todos"



Ing. Oscar Argueta Hernández  
Director Unidad de EPS

/ra



Universidad San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala 2 de marzo de 2023

Ingeniero  
**Carlos Gustavo Alonzo**  
Director de la Escuela de Ingeniería  
En Ciencias y Sistemas

Respetable Ingeniero Alonzo:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación-EPS de la estudiante **SANDRA EUNICE JIMENEZ RODAS** carné **201408470** y CUI **2976 03787 0101**, titulado: **“SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA”** y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,

A handwritten signature in black ink, appearing to read "C. Azurdia", written over a horizontal line.



**Ing. Carlos Alfredo Azurdia**  
Coordinador de Privados  
y Revisión de Trabajos de Graduación

UNIVERSIDAD DE SAN CARLOS  
DE GUATEMALA



FACULTAD DE INGENIERÍA

LNG.DIRECTOR.168.EICCSS.2023

El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador de área y la aprobación del área de lingüística del trabajo de graduación titulado: **SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, presentado por: **Sandra Eunice Jimenez Rodas**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería.

“ID Y ENSEÑAD A TODOS”



Ing. Carlos Gustavo Alonzo

Msc. Ing. Carlos Gustavo Alonzo

Director

Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, julio de 2023



LNG.DECANATO.OI.571.2023

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **SISTEMA DE INTELIGENCIA DE NEGOCIOS PARA EL MONITOREO DEL RENDIMIENTO ESTUDIANTIL PARA LA UNIDAD DE ANÁLISIS DE DATOS DE LA ESCUELA DE CIENCIAS DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, presentado por: **Sandra Eunice Jimenez Rodas**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Ing. José Francisco Gómez Rivera★

Decano a.i.

Guatemala, julio de 2023

AACE/gaoc

## **ACTO QUE DEDICO A:**

- Dios** Por demostrarme su infinito amor en mi vida, brindarme fortaleza para culminar esta meta y permitirme compartirla con quienes amo.
- Mis padres** Leonel Jiménez y Sandra Rodas, por su amor incondicional, sus consejos, creer en mí y por todas sus enseñanzas que me han guiado a lo largo de mi vida.
- Mis hermanos** Mario, Romelia y Clara Jiménez, por sus consejos y animarme ante toda circunstancia, pero sobre todo por ser mi modelo a seguir.
- Mi novio** Josué Pérez, por quererme, cuidarme y apoyarme en todo el recorrido para conseguir este logro y por inspirarme a ser mejor persona.
- Mis amigos** Jessica Martínez, Alejandro Rodríguez, Breyner Cortez y Gabriela Foronda, por ser incondicionales, por su cariño y apoyo, por sus consejos y cada momento compartido.

## **AGRADECIMIENTOS A:**

<b>Universidad de San Carlos de Guatemala</b>	Por ser mi casa de estudio y darme la oportunidad de permitirme ser parte de esta gran institución.
<b>Facultad de Ingeniería</b>	Por aceptarme como miembro de esta gran facultad y permitirme formarme en ella.
<b>Mis amigos</b>	Por llegar en distintas etapas de mi vida y compartirme su cariño, apoyo y marcarme de diferente manera.
<b>Mi familia</b>	Por llenarme de cariño y apoyo en todo momento y por sus valiosas enseñanzas.

## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES .....	VII
LISTA DE SÍMBOLOS .....	XIII
GLOSARIO .....	XV
RESUMEN .....	XIX
OBJETIVOS.....	XXI
INTRODUCCIÓN.....	XXIII
1. FASE DE INVESTIGACIÓN .....	1
1.1. Descripción de la institución .....	1
1.1.1. Reseña histórica .....	2
1.1.2. Misión.....	4
1.1.3. Visión .....	4
1.2. Descripción de las necesidades .....	4
1.2.1. Estandarización de datos en archivos.....	5
1.2.2. Diseño y creación de <i>Data Warehouse</i> .....	6
1.2.3. Proceso de extracción, transformación y carga (ETL) .....	7
1.2.4. Solución de inteligencia de negocios .....	7
1.2.5. Creación de cuadros de mando para visualización de datos .....	8
1.2.6. Administración de usuarios.....	9
1.2.6.1. Usuario final.....	9
1.2.6.2. Administrador .....	9
1.3. Priorización de las necesidades .....	9

2.	FASE TÉCNICO-PROFESIONAL .....	11
2.1.	Descripción del proyecto .....	11
2.2.	Diagnóstico de la situación actual .....	12
2.2.1.	Análisis FODA .....	12
2.3.	Investigación preliminar para la solución del proyecto.....	15
2.3.1.	Inteligencia de negocios .....	15
2.3.2.	Proceso ETL .....	15
2.3.2.1.	Fase de extracción .....	16
2.3.2.2.	Fase de transformación.....	16
2.3.2.3.	Fase de carga .....	16
2.3.3.	<i>Data Warehouse</i> .....	17
2.3.4.	Datamart .....	17
2.3.5.	Cubo multidimensional OLAP.....	17
2.3.5.1.	Métricas .....	18
2.3.5.2.	Dimensión.....	18
2.3.5.3.	Jerarquía .....	18
2.3.6.	Cuadros de mando .....	18
2.3.7.	Diseño.....	19
2.3.7.1.	Herramientas.....	20
2.4.	Presentación de la solución del proyecto.....	21
2.4.1.	Estandarización de datos en archivos.....	21
2.4.2.	Diseño de <i>Data Warehouse</i> .....	23
2.4.2.1.	Diseño de <i>Data Warehouse</i> para la Facultad de Ingeniería.....	24
2.4.2.1.1.	Diseño del modelo conceptual para la Facultad de Ingeniería..	24

2.4.2.1.2.	Diseño del modelo lógico para la Facultad de Ingeniería.....	25
2.4.2.2.	Diseño de <i>Data Warehouse</i> para el Programa Académico Preparatorio.....	31
2.4.2.2.1.	Diseño de modelo conceptual para el Programa Académico Preparatorio .....	31
2.4.2.2.2.	Diseño del modelo lógico para el Programa Académico Preparatorio .....	32
2.4.2.3.	Diseño de <i>Data Warehouse</i> para las pruebas básicas .....	36
2.4.2.3.1.	Diseño del modelo conceptual para las pruebas básicas.....	36
2.4.2.3.2.	Diseño del modelo lógico para las pruebas básicas.....	37
2.4.2.4.	Diseño de <i>Data Warehouse</i> para la prueba específica .....	41
2.4.2.4.1.	Diseño del modelo conceptual para la prueba específica.....	42
2.4.2.4.2.	Diseño del modelo lógico para la prueba específica.....	43

2.4.3.	Proceso ETL.....	47
2.4.3.1.	Proceso de ETL para las notas de los estudiantes de la Facultad de Ingeniería.....	48
2.4.3.2.	Proceso de ETL para las notas de los estudiantes del curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio.....	51
2.4.3.3.	Proceso de ETL para las notas de los estudiantes que aprobaron las pruebas básicas de física y lenguaje ...	52
2.4.3.4.	Proceso de ETL para las notas de los estudiantes que realizaron la prueba específica de matemática.....	54
2.4.4.	Cubo multidimensional OLAP.....	56
2.4.4.1.	Cubo multidimensional para la Facultad de Ingeniería.....	57
2.4.4.2.	Cubo multidimensional para el Programa Académico Preparatorio.....	61
2.4.4.3.	Cubo multidimensional para las pruebas básicas .....	64
2.4.4.4.	Cubo multidimensional para la prueba específica .....	67
2.4.5.	Cuadros de mando .....	69
2.4.5.1.	Cuadros de mando para la Facultad de Ingeniería .....	70
2.4.5.2.	Cuadros de mando para el Programa Académico Preparatorio.....	74

2.4.5.3.	Cuadros de mando para las pruebas básicas.....	76
2.4.5.4.	Cuadros de mando para la prueba específica .....	79
2.4.6.	Sitio web .....	82
2.4.6.1.	Usuario final.....	83
2.4.6.2.	Administrador .....	83
2.4.6.3.	Incrustar cuadros de mando a sitio web .....	83
2.4.6.4.	Administración de usuarios .....	84
2.4.6.5.	Carga de archivos .....	85
2.5.	Costos del proyecto .....	86
2.6.	Beneficios del proyecto.....	87
3.	FASE DE ENSEÑANZA-APRENDIZAJE .....	89
3.1.	Material elaborado .....	89
3.2.	Manual de instalación y configuración.....	89
3.3.	Manual técnico .....	89
3.4.	Manual de usuario .....	90
	CONCLUSIONES .....	91
	RECOMENDACIONES.....	93
	REFERENCIAS .....	95



# ÍNDICE DE ILUSTRACIONES

## FIGURAS

1.	Diseño del sistema.....	19
2.	Modelo conceptual para la Facultad de Ingeniería .....	25
3.	Modelo lógico de la dimensión tiempo de la Facultad de Ingeniería.....	26
4.	Modelo lógico de dimensiones de la Facultad de Ingeniería .....	27
5.	Modelo lógico de tabla de hechos para las notas de la Facultad de Ingeniería .....	29
6.	Modelo lógico de <i>Data Warehouse</i> para la Facultad de Ingeniería .....	30
7.	Modelo conceptual del Programa Académico Preparatorio .....	31
8.	Modelo lógico de la dimensión tiempo del Programa Académico Preparatorio .....	33
9.	Modelo lógico de las dimensiones del Programa Académico Preparatorio .....	33
10.	Modelo lógico de la tabla de hechos para las notas del Programa Académico Preparatorio .....	35
11.	Modelo lógico del <i>Data Warehouse</i> para el Programa Académico Preparatorio .....	35
12.	Modelo conceptual de las pruebas básicas.....	36
13.	Modelo lógico de la dimensión tiempo de las pruebas básicas .....	38
14.	Modelo lógico de las dimensiones de las pruebas básicas .....	38
15.	Modelo lógico de la tabla de hechos para las notas de las pruebas básicas.....	40

16.	Modelo lógico de <i>Data Warehouse</i> para las notas de las pruebas básicas.....	41
17.	Modelo conceptual de la prueba específica.....	42
18.	Modelo lógico de la dimensión tiempo de la prueba específica.....	44
19.	Modelo lógico de las dimensiones de la prueba específica.....	44
20.	Modelo lógico de la tabla de hechos para las notas de la prueba específica.....	46
21.	Modelo lógico del <i>Data Warehouse</i> para las notas de la prueba específica.....	47
22.	Proceso de ETL para la Facultad de Ingeniería.....	49
23.	Proceso de ETL para el Programa Académico Preparatorio.....	51
24.	Proceso de ETL para las pruebas básicas.....	53
25.	Proceso de ETL para la prueba específica de matemática.....	55
26.	Cubo multidimensional para la Facultad de Ingeniería.....	57
27.	Cubo multidimensional para el Programa Académico Preparatorio.....	62
28.	Cubo multidimensional para las pruebas básicas.....	65
29.	Cubo multidimensional para las específicas.....	67
30.	Cuadro de mando para el monitoreo del rendimiento estudiantil de la Facultad de Ingeniería.....	71
31.	Cuadro de mando para el monitoreo del rendimiento estudiantil del curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio.....	74
32.	Cuadro de mando para el monitoreo del rendimiento estudiantil para las pruebas básicas de física y lenguaje.....	77
33.	Cuadro de mando para el monitoreo del rendimiento estudiantil para la prueba específica de matemática.....	79
34.	Módulo de rendimiento estudiantil.....	84
35.	Módulo de administración de usuarios.....	85
36.	Módulo de carga de archivos.....	86

## TABLAS

I.	Prioridad e impacto de las necesidades.....	10
II.	Matriz FODA del sistema de inteligencia de negocios para la Unidad de Análisis de Datos.....	14
III.	Selección de herramientas .....	20
IV.	Formato de archivos para notas de la Facultad de Ingeniería .....	21
V.	Formato de archivos para cursos de la Facultad de Ingeniería .....	21
VI.	Formato de archivos para el pensum de las redes de estudio de la Facultad de Ingeniería .....	22
VII.	Formato de archivos para notas del Programa Académico Preparatorio .....	22
VIII.	Formato de archivos para notas de las pruebas básicas de física y lenguaje .....	23
IX.	Formato de archivos para notas de la prueba específica de matemática .....	23
X.	Detalle de dimensiones en modelo lógico para la Facultad de Ingeniería.....	28
XI.	Detalle de dimensiones en modelo lógico para el Programa Académico Preparatorio .....	34
XII.	Detalle de dimensiones en modelo lógico para las pruebas básicas ...	39
XIII.	Detalle de dimensiones en modelo lógico para la prueba específica ...	45
XIV.	Componentes de <i>job</i> principal .....	49
XV.	Tareas realizadas en el <i>job</i> principal para la Facultad de Ingeniería....	50
XVI.	Tareas realizadas en el <i>job</i> principal para el Programa Académico Preparatorio .....	52
XVII.	Tareas realizadas en el <i>job</i> principal para las pruebas básicas de física y lenguaje .....	53

XVIII.	Tareas realizadas en el <i>job</i> principal para la prueba específica de matemática .....	55
XIX.	Componentes del cubo multidimensional.....	58
XX.	Dimensiones y jerarquías en cubo multidimensional para la Facultad de Ingeniería.....	59
XXI.	Métricas en cubo multidimensional para la Facultad de Ingeniería .....	59
XXII.	Miembros calculados en cubo multidimensional para la Facultad de Ingeniería.....	61
XXIII.	Dimensiones y jerarquías en cubo multidimensional para el Programa Académico Preparatorio .....	62
XXIV.	Métricas en cubo multidimensional para el Programa Académico Preparatorio .....	63
XXV.	Miembros calculados en cubo multidimensional para el Programa Académico Preparatorio .....	64
XXVI.	Dimensiones y jerarquías en cubo multidimensional para las pruebas básicas.....	65
XXVII.	Métricas en cubo multidimensional para las pruebas básicas .....	66
XXVIII.	Miembros calculados en cubo multidimensional para las pruebas básicas .....	66
XXIX.	Dimensiones y jerarquías en cubo multidimensional para la prueba específica.....	68
XXX.	Métricas en cubo multidimensional para la prueba específica .....	68
XXXI.	Miembros calculados en cubo multidimensional para las pruebas básicas .....	69
XXXII.	Reportes generales para el monitoreo del rendimiento estudiantil de la Facultad de Ingeniería .....	72
XXXIII.	Reportes por curso para el monitoreo del rendimiento estudiantil de la Facultad de Ingeniería .....	73

XXXIV.	Reportes para el monitoreo del rendimiento estudiantil del curso de Matemática para Ingeniería del Programa Académico Preparatorio ....	75
XXXV.	Reportes para el monitoreo del rendimiento estudiantil de los estudiantes aprobados de las pruebas básicas de física y lenguaje ....	78
XXXVI.	Reportes generales para el monitoreo del rendimiento estudiantil de los estudiantes que realizaron la prueba específica de matemática.....	80
XXXVII.	Reportes por oportunidad para el monitoreo del rendimiento estudiantil de los estudiantes que realizaron la prueba específica de matemática .....	81
XXXVIII.	Detalle de costos del proyecto.....	86



## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
<b>Q</b>	Quetzal (moneda guatemalteca)



## GLOSARIO

<b>Arquitectura de software</b>	Diseño que define la estructura del sistema en su nivel más alto.
<b><i>Backend</i></b>	Aplicación del sistema que realiza la parte lógica.
<b>Base de datos</b>	Colección de datos organizados y almacenados para su posterior uso.
<b>BI</b>	Inteligencia de negocios.
<b>Canva</b>	Aplicación web para la creación de diseños gráficos.
<b>Código abierto</b>	Programa desarrollado para su uso público que permite la colaboración abierta con múltiples programadores.
<b>Confidencialidad</b>	Limita el acceso a la información por medio de la autenticación.
<b><i>Dashboard</i></b>	Cuadro de mando compuesto por múltiples reportes que muestran indicadores que aportan valor para los interesados.
<b><i>Data Warehouse</i></b>	Almacén de datos de una organización diseñado para analizar los datos de una mejor manera.

<b>Disponibilidad</b>	Garantiza que la información solo será visible por personas autorizadas.
<b>ETL</b>	Proceso de extracción, transformación y carga de datos.
<b>Filtro</b>	Procesa un flujo de datos por medio de un parámetro con el cual se obtienen resultados más específicos.
<b><i>Framework</i></b>	Estructura de trabajo para desarrollar un sistema en menos tiempo utilizando buenas prácticas.
<b><i>Iframe</i></b>	Etiqueta de HTML que permite incrustar un documento externo al sitio web.
<b>Integridad</b>	Asegura que la información es precisa y que los datos no son alterados.
<b><i>Job</i></b>	Coordina el flujo de las actividades de un proceso de ETL.
<b>MySQL</b>	Sistema de código abierto para la gestión de bases de datos relacionales.
<b>OLAP</b>	Procesamiento analítico en línea, útil en la inteligencia de negocios para agilizar las consultas de grandes cantidades de datos.

<b>Pentaho</b>	Plataforma de código abierto que cuenta con herramientas para la solución de inteligencia de negocios.
<b><i>Plugin</i></b>	Programa complementario que brinda más funciones en una aplicación.
<b>XML</b>	Lenguaje de marcado extensible para el intercambio de información estructurada.



## RESUMEN

Implementar la inteligencia de negocios en una institución brinda la oportunidad de agrupar datos significativos y convertirlos en conocimiento, esto permite a las organizaciones analizar y comprender fácilmente el comportamiento de sus datos. Se implementó una inteligencia de negocios para la Unidad de Análisis de Datos de la Escuela de Ciencias de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, ya que cada año recibe las notas de los estudiantes en archivos Excel, por lo que les resulta complejo hacer un análisis y llevar un seguimiento de los datos históricamente.

La solución realizada tiene la capacidad de extraer los datos de los archivos recibidos, seleccionar y transformar los datos que aportan valor, así como cargar los datos a un almacén de datos, el cual permite guardar grandes cantidades de estos, lleva un registro histórico y está diseñado estratégicamente para procesar y analizar los datos fácilmente en el menor tiempo posible.

Sobre la información almacenada se obtuvieron indicadores útiles para monitorear el desempeño de los estudiantes y se crearon cuadros de mando compuestos por reportes que muestran la información de manera ordenada. Dichos cuadros de mando únicamente pueden ser accedidos por las personas que forman parte de la institución a través de un sitio web, en donde podrán visualizar los datos de mejor manera y tomar decisiones con base en los datos.



# OBJETIVOS

## General

Elaborar un sistema de inteligencia de negocios para el monitoreo del rendimiento académico de los estudiantes de la Escuela de Ciencias y de la Facultad de Ingeniería, de los estudiantes del Programa Académico Preparatorio y de los aspirantes a ingresar a la Facultad de Ingeniería que realizan las pruebas básicas y específicas de la Universidad de San Carlos de Guatemala.

## Específicos

1. Realizar un proceso de extracción, transformación y carga sobre los archivos con las notas de los estudiantes de la Facultad de Ingeniería, del Programa Académico Preparatorio y de los resultados de quienes aprobaron las pruebas básicas y específicas, almacenando la información en un *Data Warehouse*.
2. Implementar una solución de inteligencia de negocios para el monitoreo del rendimiento estudiantil sobre las notas de los estudiantes de los cursos propios de la Facultad de Ingeniería, de las notas de los estudiantes del curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio y de los resultados aprobados de las pruebas de conocimientos básicos de lenguaje y física y la prueba específica de Matemática para Ingeniería.

3. Crear una aplicación web para visualizar el rendimiento académico a través de cuadros de mando, en donde se permita aplicar filtros sobre la información de las notas obtenidas de los estudiantes de la Facultad de Ingeniería, del Programa Académico Preparatorio y de los resultados de las pruebas básicas y específicas.

## INTRODUCCIÓN

La Unidad de Análisis de Datos de la Escuela de Ciencias de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala recibe cada año archivos Excel con los datos de las notas obtenidas por los estudiantes de la Facultad de Ingeniería, de quienes aprueban las pruebas de conocimiento básico de física y lenguaje, de quienes realizan la prueba específica de matemática para ingresar a la Facultad de Ingeniería y las notas del curso de Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio.

La Unidad de Análisis de Datos posee una gran cantidad de datos, la cual incrementa cada año, haciendo difícil el análisis y seguimiento de los datos de los estudiantes de manera manual. La institución actualmente no cuenta con un sistema que facilite el análisis de los datos, por lo que solicita un sistema que procese la información, garantizando la privacidad de los datos y que permita monitorear el rendimiento de los estudiantes únicamente a personas autorizadas.

Ante las necesidades expuestas, el presente trabajo presenta una solución de inteligencia de negocios que transformó los datos en conocimiento sobre el rendimiento de los estudiantes. Para ello se hizo uso de herramientas de código abierto, las cuales facilitaron los procesos para crear una inteligencia de negocios. A los interesados se les presentó un sitio web con cuadros de mando para visualizar los datos de manera fácil de comprender para la toma de decisiones.



# **1. FASE DE INVESTIGACIÓN**

## **1.1. Descripción de la institución**

Según el portal oficial de Internet de la Facultad de Ingeniería de la USAC (2022):

La Facultad de Ingeniería de la Universidad de San Carlos de Guatemala tiene como función formar profesionales de prestigio que aporten a la nación. Esta es una unidad académica que dispone de doce programas de pregrado y ofrece una escuela de postgrado con carácter regional centroamericano, además del Centro de Investigaciones de Ingeniería.

En la actualidad cuenta con ocho escuelas de nivel pregrado, donde se encuentra la Escuela de Ciencias, la cual atiende la etapa básica para las diferentes carreras de ingeniería. Dentro de la Escuela de Ciencias se encuentra la Unidad de Análisis de Datos, la cual vela por el mejoramiento de la calidad de la formación matemática y ciencias básicas impartidas a los estudiantes de las carreras de la Facultad de Ingeniería.

Como parte del proceso de ingreso a la prestigiosa Facultad de Ingeniería, es necesario que los aspirantes realicen pruebas de conocimientos básicos y específicos. Quienes reprobaban pueden optar al Programa Académico Preparatorio, el cual es un programa de nivelación que cuenta con cursos preparatorios para ingresar a la Facultad con el objetivo de nivelar los conocimientos en las materias.

### **1.1.1. Reseña histórica**

Con base en la información proporcionada por la página de Internet de la Facultad de Ingeniería de la USAC (2022):

La Facultad de Ingeniería en la Universidad de San Carlos de Guatemala surgió de una serie de sucesos dentro de la Universidad. En el año 1834 se creó la Academia de Ciencias, la cual fue pionera de los estudios de ciencias tanto dentro de la Universidad como en el país en general. En la Academia de Ciencias se impartió la enseñanza de álgebra, geometría, trigonometría y física. Cabe destacar que José Batres Montúfar se graduó con el título de ingeniero agrimensor de esta academia. En el año 1840 dejó de funcionar.

Después de la Revolución de Guatemala en 1871 se dio un giro en la enseñanza técnica superior, en 1873 se fundó la Escuela Politécnica para formar ingenieros militares, topógrafos, telégrafos y oficiales militares. En 1879 se creó la Escuela de Ingeniería por decreto del gobierno y en 1882 se tituló como Facultad de Ingeniería, separándose de la Escuela Politécnica.

Debido a la economía, en 1894 la Facultad de Ingeniería nuevamente se unió a la Escuela Politécnica, marcando paso a un periodo de inestabilidad para la Facultad. Durante este periodo estuvo ubicada en distintas localidades y fue hasta 1920 que se instaló en un edificio ubicado frente al parque de Jocotenango, en la zona 2 de la Ciudad de Guatemala. En ese año la Facultad reanudó sus labores, disponiendo únicamente de la carrera para ingeniero topógrafo. Fue hasta 1930 que se agregó la carrera de ingeniero civil, marcando paso a la época moderna.

En 1967 se integró a la Facultad de Ingeniería la Escuela de Ingeniería Química y se creó la Escuela de Ingeniería Mecánica Industrial, la cual tenía las carreras de Ingeniería Industrial, Ingeniería Mecánica e Ingeniería Mecánica Industrial. En el siguiente año se creó la Escuela de Ingeniería Mecánica Eléctrica, a cargo de las carreras de Ingeniería Eléctrica e Ingeniería Mecánica Eléctrica. En 1970 se crea la Escuela de Ciencias y Sistemas con grado de licenciatura.

En 1971 se llevó a cabo el plan de reestructuración de la Facultad de Ingeniería, el cual incitaba a los estudiantes de la Facultad a participar en el desarrollo del país. En 1974 se creó la Unidad de Ejercicio Profesional Supervisado para todas las carreras de la Facultad de Ingeniería y en 1976 se creó la Escuela de Ciencias, con el propósito de atender la etapa común de las carreras de ingeniería. Dentro de la Escuela de Ciencias surgieron las carreras de Licenciatura en Matemática Aplicada y Licenciatura en Física Aplicada en 1980.

En 1989, gracias al avance tecnológico, se creó la carrera de Ingeniería Electrónica en la Escuela de Ingeniería Mecánica Eléctrica. En 1999 inició un proceso de exámenes de ubicación para los estudiantes de primer ingreso y un programa de nivelación con cursos para nivelar a los aspirantes a ingresar a la Facultad de Ingeniería, y en el mismo año se agregaron cursos de Inglés Técnico para las carreras de ingeniería.

En 2004 se dio la opción a los estudiantes de todas las carreras de obtener un Diplomado en Administración de Empresas. En 2005 se creó la carrera de Ingeniería Ambiental con grado en licenciatura.

La Facultad de Ingeniería se ubica en la Universidad de San Carlos de Guatemala, en la zona 12 de la ciudad de Guatemala.

### **1.1.2. Misión**

Formar profesionales en las distintas áreas de la ingeniería que, a través de la aplicación de la ciencia y la tecnología, conscientes de la realidad nacional y regional, y comprometidos con nuestras sociedades, sean capaces de generar soluciones que se adapten a los desafíos del desarrollo sostenible y los retos del contexto global. (Facultad de Ingeniería de la USAC, 2022)

### **1.1.3. Visión**

Ser una institución académica con incidencia en la solución de la problemática nacional; formamos profesionales en las distintas áreas de la ingeniería, con sólidos conceptos científicos, tecnológicos, éticos y sociales, fundamentados en la investigación y promoción de procesos innovadores orientados hacia la excelencia profesional. (Facultad de Ingeniería de la USAC, 2022)

## **1.2. Descripción de las necesidades**

La Unidad de Análisis de Datos de la Escuela de Ciencias de la Facultad de Ingeniería recibe la siguiente información en formato Excel:

- Notas obtenidas por los estudiantes en los cursos propios de la Facultad de Ingeniería durante el primer y segundo semestre y en cursos de vacaciones servidos en los meses de junio y diciembre de cada año.

- Notas obtenidas por los estudiantes asignados a la Facultad de Ingeniería con resultado satisfactorio en las pruebas de conocimientos básicos de lenguaje y física.
- Notas obtenidas por las personas que realizan la prueba específica de Matemática para Ingeniería.
- Notas obtenidas por los estudiantes del curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio.

Con la información proporcionada, la Unidad de Análisis de Datos solicita la creación de un sistema que le permita identificar el rendimiento académico de los estudiantes. Para ello se plantearon los siguientes productos y las soluciones.

### **1.2.1. Estandarización de datos en archivos**

Es necesario que se estandarice el formato de los archivos recibidos sobre las notas de los estudiantes para que exista un orden sobre los datos, ya que mantener la información estandarizada supone una gran ventaja a nivel empresarial. Estandarizar los datos permite organizar conjuntos de información compleja y facilitar el análisis de la información.

Con base en la información proporcionada por la página de Internet Ayuware (2022) estandarizar los datos también permite a los interesados tener archivos con información que será útil y descarta los datos que no se desean o que pueden contener información sensible para tratar, como datos personales de los estudiantes.

- Solución

Elaborar estándares sobre los archivos recibidos por la institución, para ello se debe planificar junto con los encargados aquellos datos que serán útiles de almacenar y cuáles por seguridad de la información no deben ser tomados en cuenta. Al definir el estándar de los datos se desarrollará un programa para procesar los archivos y unificar la información en uno solo con el formato establecido.

### **1.2.2. Diseño y creación de *Data Warehouse***

Es necesaria la creación de un *Data Warehouse* para almacenar los resultados de los estudiantes y eliminar los datos que no serán útiles en el análisis de la información. Un *Data Warehouse* puede ser estructurado de tal manera que posteriormente se permita analizar de mejor manera la información y así tener una mejor comprensión sobre los datos.

- Solución

Diseñar un *Data Warehouse* conforme a las necesidades de la institución, se debe definir el esquema a utilizar, el cual debe ser capaz de almacenar la información de tal manera que sea fácil analizarla. Un *Data Warehouse* permitirá integrar la información de los resultados de los años a través del tiempo, así se guardarán datos históricos y se almacenarán grandes cantidades de información.

### **1.2.3. Proceso de extracción, transformación y carga (ETL)**

Es necesario realizar un proceso de extracción, transformación y carga sobre los archivos planos recibidos por la institución, ya que, antes de ser cargados al *Data Warehouse* deben ser analizados para verificar que cumplan con el formato definido, y también es necesario que se realice una limpieza de los datos y sean transformados para que tengan sentido y de esta manera puedan ser cargados y utilizados para su análisis.

- Solución

Es indispensable que se verifiquen los datos y se realice un proceso de limpieza y transformación antes de que sean cargados al *Data Warehouse* para asegurar que solo se almacenen datos útiles para el análisis. Por lo tanto, se desarrollará un proyecto con una herramienta de software libre para realizar el proceso de extracción, transformación y carga sobre los archivos planos recibidos.

### **1.2.4. Solución de inteligencia de negocios**

Es necesario para la institución contar con un sistema que permita el monitoreo del rendimiento académico de los estudiantes, por lo que la creación de un sistema de inteligencia de negocios es primordial para que se formulen y respondan preguntas acerca de los datos. Esto también favorecerá la toma de decisiones y facilitará la creación de los reportes, ya que se podrán visualizar datos actuales e históricos sobre los resultados de los estudiantes.

- Solución

Crear una solución de inteligencia de negocios permitirá a la institución llevar un control sobre los datos actuales e históricos de los resultados de los estudiantes, así los interesados podrán monitorear el rendimiento de los estudiantes para verificar si se cumple con las expectativas y, a partir de los datos, encontrar deficiencias y tomar decisiones para obtener mejores resultados.

#### **1.2.5. Creación de cuadros de mando para visualización de datos**

Es necesario mostrar la información de una manera clara y ordenada, para permitir la toma de decisiones y el monitoreo rápido de parte de la institución, por lo que estructurar la información en cuadros de mando permitirá mostrar los datos relevantes que son necesarios para el monitoreo del rendimiento en varios reportes agrupados de manera entendible.

- Solución

Con el fin de mostrar la información de manera segura, se desarrollará un sitio web en donde únicamente los interesados de la institución podrán visualizar los datos a través de cuadros de mando. Por lo tanto, se realizará un análisis sobre los reportes necesarios a mostrar para medir el rendimiento estudiantil.

### **1.2.6. Administración de usuarios**

Se necesita que únicamente personas que pertenecen a la institución puedan acceder al sitio web, esto para que la información no sea expuesta abiertamente al público. Se definirán los usuarios y qué beneficios obtendrá cada uno dentro del sitio web.

#### **1.2.6.1. Usuario final**

Este usuario tendrá permitido visualizar los cuadros de mando del rendimiento académico, de los resultados de los estudiantes de la Facultad de Ingeniería, del curso de Matemática para Ingeniería del Programa Académico Preparatorio y de los resultados de los aspirantes a ingresar a la Facultad de Ingeniería que realizan las pruebas básicas de física y lenguaje y que realizan la prueba específica de matemática.

#### **1.2.6.2. Administrador**

Este usuario es quien contará con mayores privilegios que los de un usuario final. El administrador tendrá acceso a subir al sistema los archivos brindados con los datos de los resultados de los estudiantes, también tendrá permitido agregar, modificar, listar y eliminar usuarios.

### **1.3. Priorización de las necesidades**

- Definición de prioridad
  - 1 a 3: prioridad baja
  - 4 a 7: prioridad media

- 8 a 10: prioridad alta
- Definición de impacto, con base en Proactivanet (2022):
  - Alto: el servicio se ve afectado de manera severa impidiendo su uso y afectando a actividades críticas del sistema.
  - Medio: el servicio se ve afectado impidiendo su uso, pero no afectando a actividades críticas del sistema.
  - Bajo: el servicio se ve afectado, pero no impide su uso.

Tabla I. **Prioridad e impacto de las necesidades**

<b>Necesidad</b>	<b>Prioridad</b>	<b>Impacto</b>
<b>Estandarización de datos en los archivos</b>	10	Alto
<b>Diseño y creación de Data Warehouse</b>	8	Alto
<b>Proceso de extracción, transformación y carga (ETL)</b>	8	Alto
<b>Solución de inteligencia de negocios</b>	7	Alto
<b>Creación de cuadros de mando para visualización de datos</b>	6	Medio
<b>Administración de usuarios</b>	3	Bajo

Fuente: elaboración propia.

## 2. FASE TÉCNICO-PROFESIONAL

### 2.1. Descripción del proyecto

La Unidad de Análisis de Datos de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala recibe información sobre las notas de los estudiantes en los cursos por carrera de la Facultad de Ingeniería, también recibe las notas de los estudiantes del curso de Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio y las notas de los aspirantes a ingresar a la Facultad de Ingeniería que realizan las pruebas de conocimientos básicos y específicos.

Se espera realizar un análisis sobre estos datos para examinar el rendimiento de los estudiantes que aspiran a ingresar y de quienes ya forman parte de la Facultad de Ingeniería, para ello se plantea realizar un sistema de inteligencia de negocios para convertir la información que se recibe en conocimiento que le permita a la Unidad de Análisis de Datos monitorear el rendimiento estudiantil y les ayude a identificar el comportamiento de los estudiantes en los cursos de la Facultad de Ingeniería.

La información recibida sobre los estudiantes de la Facultad de Ingeniería y de quienes aspiran a ingresar será almacenada en un *Data Warehouse* por su fácil acceso y administración, para ello se realizará un proceso de ETL, ya que dicha información es recibida en archivos planos. Los datos se extraerán de los archivos, se transformarán validando la integridad de los datos y seguido de la limpieza de los datos se cargarán en un *Data*

*Warehouse* con el fin de almacenarlos de manera ordenada para favorecer la comprensión y el análisis.

La información almacenada será analizada por un cubo de información, es decir, una estructura multidimensional, la cual permitirá buscar los datos con mayor rapidez, independientemente de los datos en el cubo o de la complejidad de la búsqueda. A esta estructura comúnmente se le llama cubo multidimensional OLAP (Online Analytic Processing) y los datos dentro del cubo son categorizados por dimensiones, los cuales proveen una descripción de las medidas numéricas que servirán para el análisis del rendimiento de los estudiantes.

Estos datos serán visualizados a través de herramientas que permitirán navegar en ellos, se crearán cuadros de mando que mostrarán los reportes y la información relevante para monitorear el rendimiento de los estudiantes.

## **2.2. Diagnóstico de la situación actual**

Mediante reuniones con los interesados se pudo identificar los factores que influyen en el desarrollo del proyecto y por medio de ello se realizó un análisis FODA para dar a conocer el diagnóstico de la situación actual.

### **2.2.1. Análisis FODA**

Se realizó un análisis sobre los factores internos y externos que tienen mayor impacto con respecto a la elaboración del proyecto para la institución, siendo los principales hallazgos clasificados en las categorías del FODA los siguientes:

- Análisis interno
  - Fortalezas
    - F1: transformación de datos sobre notas reales de los estudiantes para el monitoreo del rendimiento académico.
    - F2: visualización de datos históricos sobre las notas de los estudiantes precisando el menor tiempo posible de procesamiento.
    - F3: visualización de datos ordenados para la toma de decisiones.
  - Debilidades
    - D1: el uso de herramientas de código libre no cuenta con soporte técnico en caso de fallos.
- Análisis externo
  - Oportunidades
    - O1: toma de decisiones sobre mejoras de aprendizaje basándose en los cuadros de mando.
    - O2: incentivar el éxito de los estudiantes apoyando su aprendizaje.
    - O3: identificar si existen factores que influyen positiva o negativamente en el rendimiento de los estudiantes.
  - Amenazas
    - A1: pandemia COVID-19.
    - A2: centro de cálculo no pueda compartir datos solicitados.

- A3: centro de cálculo no envíe el formato solicitado de los archivos con las notas de los cursos de la Facultad de Ingeniería.
- A4: Escuela de Ciencias no envíe el formato solicitado de los archivos con las notas del curso de Matemática para Ingeniería del Programa Académico Preparatorio.
- A5: Sistema de Ubicación y Nivelación no envíe el formato solicitado de los archivos con las notas de las pruebas básicas y específicas para la Facultad de Ingeniería.

A continuación, en la tabla II, se presenta la matriz FODA realizada:

Tabla II. **Matriz FODA del sistema de inteligencia de negocios para la Unidad de Análisis de Datos**

<b>Fortalezas</b>	<b>Debilidades</b>
<ul style="list-style-type: none"> <li>• F1: transformación de datos para el monitoreo del rendimiento académico.</li> <li>• F2: visualización de datos históricos en el menor tiempo posible de procesamiento.</li> <li>• F3: visualización de datos ordenados para la toma de decisiones.</li> </ul>	<ul style="list-style-type: none"> <li>• D1: el uso de herramientas de código libre no cuenta con soporte técnico.</li> </ul>
<b>Oportunidades</b>	<b>Amenazas</b>
<ul style="list-style-type: none"> <li>• O1: toma de decisiones sobre mejoras de aprendizaje.</li> <li>• O2: incentivar el éxito de los estudiantes.</li> <li>• O3: identificar si existen factores que influyen en el rendimiento de los estudiantes.</li> </ul>	<ul style="list-style-type: none"> <li>• A1: pandemia COVID-19.</li> <li>• A2: centro de cálculo no pueda compartir datos solicitados.</li> <li>• A3: centro de cálculo no envíe el formato solicitado de los archivos.</li> <li>• A4: Escuela de Ciencias no envíe el formato solicitado de los archivos.</li> <li>• A5: Sistema de Ubicación y Nivelación no envíe el formato solicitado de los archivos.</li> </ul>

Fuente: elaboración propia.

## **2.3. Investigación preliminar para la solución del proyecto**

Se desarrolla este tema en los siguientes incisos:

### **2.3.1. Inteligencia de negocios**

La inteligencia de negocios o BI es un conjunto de metodologías y tecnologías que permiten realizar dos procesos importantes, los cuales son transformar datos en información y transformar esa información en conocimiento para la toma de decisiones basadas en datos. Tiene la capacidad de mostrar datos actuales e históricos y los analistas pueden aprovechar esta información para conocer y tomar en cuenta el comportamiento y rendimiento de los datos y para detectar las tendencias de ellos.

La inteligencia de negocios ayuda a las organizaciones que trazan objetivos y realizan preguntas, las cuales pueden ser resueltas realizando un análisis sobre sus datos, para ello se recopila la información necesaria, se procesa y se transforma en conocimiento y los usuarios de la organización pueden acceder a ellos de manera visual para analizarlos y responder a sus preguntas (Tableau, s.f).

La inteligencia de negocios ha evolucionado a través de los años, incluyendo procesos importantes para ayudar a mejorar el rendimiento. En este trabajo de Ejercicio Profesional Supervisado se utilizaron los siguientes.

### **2.3.2. Proceso ETL**

El proceso ETL (*Extract, Transform, Load*) que significa extraer, transformar y cargar, permite mover datos de múltiples fuentes para

almacenarlos en un lugar como una base de datos, *Datamart* o *Data Warehouse*. Como su nombre lo indica, este proceso consta de tres fases donde se busca organizar, limpiar y almacenar los datos para su análisis

#### **2.3.2.1. Fase de extracción**

Esta es la primera fase del proceso de ETL y es donde se reciben los datos que se utilizarán en las siguientes dos fases. La data es recibida desde múltiples fuentes tales como base de datos, archivos planos como Excel, CVS, entre otros. Estos datos son extraídos y se analizan para verificar que cumplan con la estructura que se esperaba recibir y de no ser así se rechazan los datos. Los datos aceptados se convierten para iniciar la siguiente fase de transformación (Nida, 2020).

#### **2.3.2.2. Fase de transformación**

La segunda fase del proceso de ETL es la que más agrega valor, ya que los datos obtenidos anteriormente son limpiados, mapeados y transformados para que tengan sentido y puedan ser utilizados para su respectivo análisis. Esta fase abarca una serie de funciones, iniciando con la limpieza de los datos recuperados, donde se eliminan los duplicados y de ser necesario se corrigen valores, seguido de ello se estructuran y con ello se obtienen datos útiles. Los datos transformados continúan a la última fase para su carga (Nida, 2020).

#### **2.3.2.3. Fase de carga**

La fase de carga es la final del proceso de ETL, los datos extraídos y ya transformados son cargados en la base de datos destino, la mayoría de los casos suele ser un *Data Warehouse* (Nida, 2020).

### **2.3.3. Data Warehouse**

Un *Data Warehouse*, desde un punto de vista técnico, es considerado como el primer paso para desarrollar la solución de un sistema de inteligencia de negocios. Este es una base de datos corporativa, o bien un almacén de datos que tiene como características integrar y depurar los datos históricos y relevantes de una empresa. Su función principal es la de guardar datos necesarios dentro de una empresa para así poder utilizarlos y transformarlos en información valiosa para la empresa y los interesados.

Un *Data Warehouse* es estructurado de tal manera que posteriormente se pueda analizar de mejor manera su información y así ayudar a tener una mejor comprensión de los datos, esto a diferencia de una base de datos en donde solo se espera almacenar los mismos.

### **2.3.4. Datamart**

Esta es una base de datos departamental que, a diferencia del *Data Warehouse*, se especializa únicamente en un área de negocio. Un *Datamart* es estructurado de tal manera que se pueda tener un acceso directo a su información, permitiendo analizar de manera más óptima los datos de un área. Un *Datamart* puede ser alimentado directamente de los datos almacenados en un *Data Warehouse*, con el fin de ayudar a un área específica dentro de la empresa para tomar mejores decisiones.

### **2.3.5. Cubo multidimensional OLAP**

Un cubo multidimensional OLAP está formado por estructuras que permiten realizar un análisis sobre bases de datos relacionales de gran

volumen, permite además que el tiempo de procesamiento sea rápido y el análisis sea fácil. Los cubos son subconjuntos de los datos almacenados y sintetizados dentro de una estructura multidimensional. Comúnmente estas estructuras son utilizadas para reportes en donde los datos son categorizados por dimensiones, con las cuales se obtienen mejores consultas en comparación con una base de datos por la manera en la que está guardada la información.

#### **2.3.5.1. Métricas**

Las métricas, también conocidas como hechos o indicadores, son datos con valores numéricos, los cuales son útiles para el análisis, estos son dependientes de las dimensiones y jerarquías.

#### **2.3.5.2. Dimensión**

Una dimensión es definida como una categoría descriptiva en que las métricas del cubo son separadas para su análisis respectivo dentro de un cubo multidimensional.

#### **2.3.5.3. Jerarquía**

Una jerarquía es una relación lógica padre-hijo entre las dimensiones o atributos de estas, con las cuales se busca realizar un análisis más detallado desde el nivel más alto al más bajo.

#### **2.3.6. Cuadros de mando**

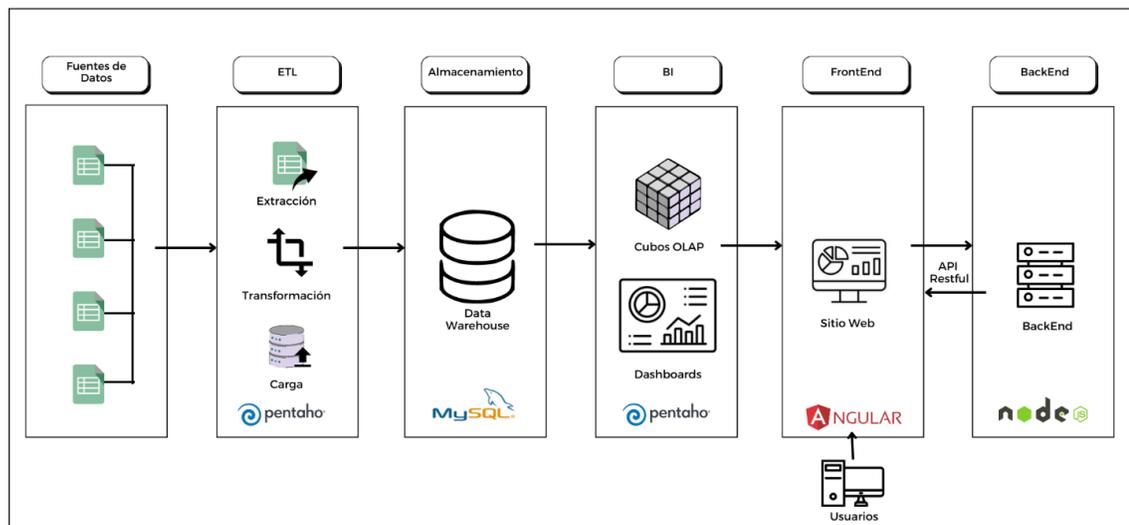
Un cuadro de mando, también conocido como *dashboard*, es una herramienta para la visualización de datos y métricas fundamentales en un

negocio o un departamento específico. Este permite visualizar los datos de manera ordenada y agradable al usuario para su fácil comprensión, con el propósito de realizar análisis sobre la información mostrada. El diseño de estos suele ser atractivo y sencillo de interpretar, usualmente es mostrado en un tablero donde se pueden visualizar distintos reportes de manera ordenada.

### 2.3.7. Diseño

Para el desarrollo de un sistema de inteligencia de negocios es necesario que se evalúe la arquitectura del sistema, así como las herramientas a utilizar que se acoplen a las necesidades de la empresa. A continuación se presenta la arquitectura del sistema:

Figura 1. Diseño del sistema



Fuente: elaboración propia.

### 2.3.7.1. Herramientas

Para el desarrollo del proyecto se evaluaron las herramientas de código libre a utilizar dependiendo de cada área del proceso de desarrollo. Se presentan en la tabla de la siguiente página:

Tabla III. Selección de herramientas

Proceso	Herramienta	Descripción
<b>Data Warehouse</b>	MySQL	Es la base de datos para soluciones de <i>Data Warehouse</i> y de inteligencia de negocios, la cual resulta efectiva con los resultados y la información almacenada.
<b>ETL</b>	<i>Pentaho Data Integration</i>	Es una herramienta de código abierto que permite que se utilicen técnicas de ETL, además ofrece datos analíticos precisos y elimina las complejidades, la cual la hace sencilla de utilizar y permite ahorrarse tiempo y esfuerzo.
<b>Cubo OLAP</b>	<i>Pentaho Schema Workbench</i>	Es una herramienta de código abierto que permite llevar la información de los <i>Data Warehouse</i> a cubos OLAP.
<b>Dashboards</b>	<i>Pentaho Business Analytics</i>	Es una herramienta de código abierto que permite ejecutar recursos como informes o cuadros de mando, aquí se permite la creación y gestión de <i>dashboards</i> .
<b>Frontend</b>	Angular	Es un <i>framework</i> utilizado para la creación de aplicaciones web, permite trabajar de manera eficiente y organizada.
<b>Backend</b>	Nodejs	Es un entorno en tiempo de ejecución para la capa de servidor basado en el lenguaje de <i>javascript</i> . Permite la creación de aplicaciones escalables, permitiendo múltiples conexiones al mismo tiempo.

Fuente: elaboración propia.

## 2.4. Presentación de la solución del proyecto

A continuación, se presenta el proceso que se llevó a cabo para el desarrollo del sistema de inteligencia de negocios para la Unidad de Análisis de Datos.

### 2.4.1. Estandarización de datos en archivos

A continuación, se presenta el estándar definido para los archivos recibidos sobre las notas de los estudiantes:

Tabla IV. **Formato de archivos para notas de la Facultad de Ingeniería**

<b>Nombre columna</b>	<b>Tipo</b>
<b>Año</b>	Entero
<b>Curso</b>	Entero
<b>nombre_curso</b>	Cadena
<b>Carnet</b>	Entero
<b>Carrera</b>	Entero
<b>Zona laboratorio</b>	Entero
<b>Zona</b>	Entero
<b>Examen final</b>	Entero

Fuente: elaboración propia.

Tabla V. **Formato de archivos para cursos de la Facultad de Ingeniería**

<b>Nombre columna</b>	<b>Tipo</b>
<b>Código</b>	Entero
<b>Área</b>	Entero
<b>Grupo</b>	Entero
<b>Créditos</b>	Entero
<b>Nombre</b>	Cadena

Fuente: elaboración propia.

Tabla VI. **Formato de archivos para el pensum de las redes de estudio de la Facultad de Ingeniería**

<b>Nombre columna</b>	<b>Tipo</b>
<b>Codigo_curso</b>	Entero
<b>Prerrequisito_1</b>	Entero
<b>Prerrequisito_2</b>	Entero
<b>Prerrequisito_3</b>	Entero
<b>Prerrequisito_4</b>	Entero
<b>Prerrequisito_creditos</b>	Entero
<b>Obligatorio</b>	Cadena
<b>Area_pensum</b>	Entero
<b>Semestre</b>	Entero
<b>Anio_pensum</b>	Entero

Fuente: elaboración propia.

Tabla VII. **Formato de archivos para notas del Programa Académico Preparatorio**

<b>Nombre columna</b>	<b>Tipo</b>
<b>Sección</b>	Cadena
<b>Nota</b>	Entero
<b>Semestre</b>	Cadena

Fuente: elaboración propia.

Tabla VIII. **Formato de archivos para notas de las pruebas básicas de física y lenguaje**

<b>Nombre columna</b>	<b>Tipo</b>
<b>Vocacional</b>	Entero
<b>Carnet</b>	Entero
<b>Código</b>	Entero
<b>Carrera</b>	Cadena
<b>Fecha asignada</b>	Fecha
<b>Resultado Leng</b>	Cadena
<b>Nota Leng</b>	Entero
<b>Resultado Física</b>	Cadena
<b>Nota Física</b>	Entero

Fuente: elaboración propia.

Tabla IX. **Formato de archivos para notas de la prueba específica de matemática**

<b>Nombre columna</b>	<b>Tipo</b>
<b>IdVocacional</b>	Entero
<b>Género</b>	Cadena
<b>Nota</b>	Entero
<b>Resultado</b>	Cadena
<b>Oportunidad</b>	Entero

Fuente: elaboración propia.

### **2.4.2. Diseño de *Data Warehouse***

Para este proyecto se realizó un *Data Warehouse* para cada entidad, con el fin de acoplar de mejor manera el modelo a las necesidades de cada una. Por cada fuente de dato se definió la estructura de su modelo lógico y, a partir de ello, se definieron las tablas de dimensiones y la tabla de hecho. Para los

*Data Warehouse* creados se empleó el esquema en estrella y/o copo de nieve implementando, el que mejor se adapta a las necesidades.

#### **2.4.2.1. Diseño de *Data Warehouse* para la Facultad de Ingeniería**

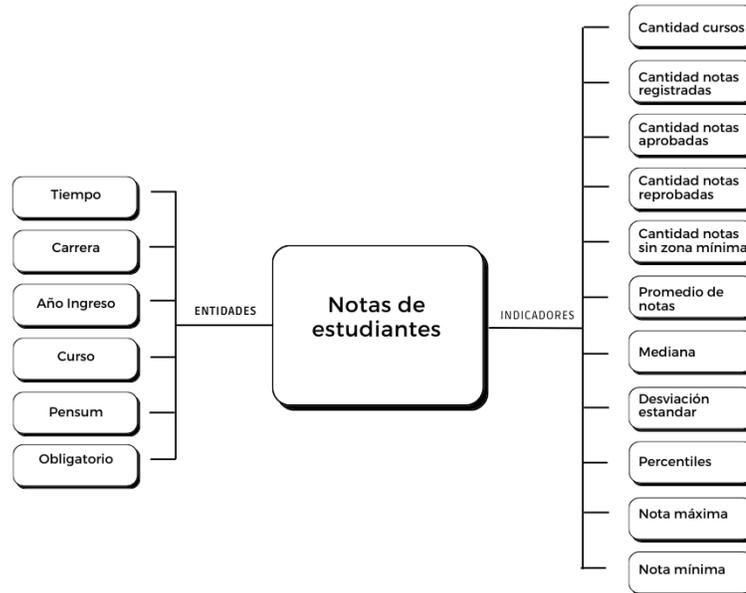
Se explica a continuación:

##### **2.4.2.1.1. Diseño del modelo conceptual para la Facultad de Ingeniería**

Para la creación del modelo conceptual se analizaron los datos proporcionados de las notas de la Facultad de Ingeniería ubicados en las tablas IV, V y VI, donde se definieron las entidades lógicas a crear y los indicadores a calcular posteriormente. Un indicador es un valor numérico de gran utilidad para analizar y responder preguntas, los indicadores pueden ser obtenidos a través de las entidades lógicas establecidas.

Un modelo conceptual para la base de datos permite visualizar de manera simple el alcance del proyecto (Bernabéu y García, 2017, p. 134). A continuación se presenta el modelo conceptual definido para las notas de la Facultad de Ingeniería:

Figura 2. **Modelo conceptual para la Facultad de Ingeniería**



Fuente: elaboración propia, empleando Canva.

En el modelo conceptual del lado izquierdo se encuentran las entidades lógicas definidas, las cuales están enlazadas a la relación que existe entre ellas, en este caso las notas de los estudiantes. Con las entidades lógicas y las notas obtenidas por los estudiantes se podrá calcular los indicadores situados del lado derecho.

#### 2.4.2.1.2. **Diseño del modelo lógico para la Facultad de Ingeniería**

El modelo lógico es la estructura del *Data Warehouse* y se crea a partir del modelo conceptual, donde las entidades definidas son las tablas de dimensiones y la relación entre ellas es la tabla de hechos.

- Tablas de dimensiones

Bernabéu y García (2017) señalan que es indispensable que en la creación de dimensiones de un *Data Warehouse* se tome en cuenta la dimensión tiempo, ya que esta permitirá llevar un registro histórico de la información almacenada. Principalmente, esta dimensión es creada primero debido a que se establece un orden en los datos (p. 52).

- Dimensión tiempo

A partir de la entidad tiempo, en el modelo conceptual se define el nombre de la dimensión, su clave principal y los campos o columnas. Para definir la jerarquía dentro de la dimensión tiempo es necesario conocer la forma en que la institución trabaja, los datos recibidos y cómo se relacionan en el tiempo, en este caso se definió la jerarquía año>semestre>periodo.

Figura 3. **Modelo lógico de la dimensión tiempo de la Facultad de Ingeniería**



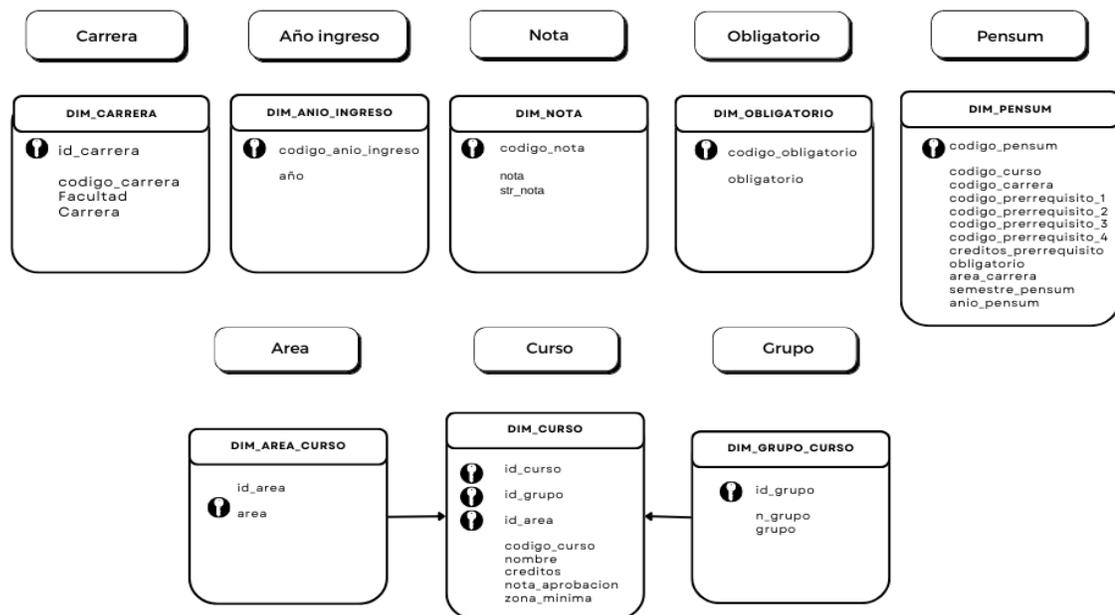
Fuente: elaboración propia, empleando Canva.

Al tener establecida la dimensión tiempo se procede a crear las dimensiones que se mantendrán casi constantes en el tiempo y que mantendrán su estructura independientemente del tiempo.

- Dimensiones independientes

A partir de las entidades en el modelo conceptual se definió para cada una el nombre de la dimensión, su clave principal, sus campos y sus jerarquías.

Figura 4. **Modelo lógico de dimensiones de la Facultad de Ingeniería**



Fuente: elaboración propia, empleando Canva.

Al establecer las dimensiones del *Data Warehouse* se procede a definir el tipo de dato para cada campo y se establece según los datos recibidos que permitirán valores nulos.

Tabla X. **Detalle de dimensiones en modelo lógico para la Facultad de Ingeniería**

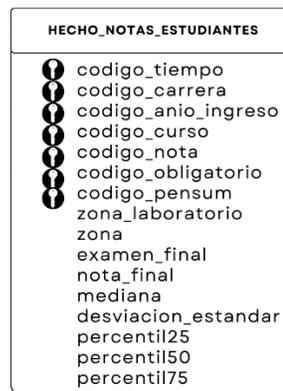
<b>Dimensión</b>	<b>Campo</b>	<b>Tipo</b>	<b>Permite nulos</b>
<b>Dim_Carrera</b>	id_carrera	Int	No
	codigo_carrera	Int	No
	facultad	Varchar(75)	No
	carrera	Varchar(75)	No
<b>Dim_anio_ingreso</b>	codigo_anio_ingreso	Int	No
	anio_ingreso	Int	No
<b>Dim_area_curso</b>	Id_area	Int	No
	área	Varchar(15)	No
<b>Dim_grupo_curso</b>	id_grupo	Int	No
	n_grupo	Int	No
	grupo	Varchar(100)	Sí
<b>Dim_curso</b>	id_curso	Int	No
	codigo_curso	Int	No
	Id_grupo	Int	Sí
	Id_area	Int	Sí
	Nombre	Varchar(100)	No
	creditos	Smallint	Sí
	nota_aprobacion	Smallint	Sí
	zona_mínima	Smallint	Sí
<b>Dim_nota</b>	codigo_nota	Int	No
	nota	Int	No
	str_nota	Varchar(10)	No
<b>Dim_Obligatorio</b>	codigo_obligatorio	Int	No
	obligatorio	Int	No
<b>Dim_Pensum</b>	codigo_pensum	Int	No
	codigo_curso	Int	No
	codigo_carrera	Int	No
	codigo_prerrequisito_1	Int	Sí
	codigo_prerrequisito_2	Int	Sí
	codigo_prerrequisito_3	Int	Sí
	codigo_prerrequisito_4	Int	Sí
	creditos_prerrequisito	Int	Sí
	obligatorio	Varchar(20)	No
	area_carrera	Int	No
	semestre_pensum	Int	No
	anio_pensum	Int	No

Fuente: elaboración propia.

- Tabla de hechos

La tabla de hechos almacena los datos llamados hechos que permitan obtener los indicadores definidos en el mapa conceptual, por lo general estos son datos cuantitativos que podrán ser agrupados o filtrados por las tablas de dimensiones del *Data Warehouse* (Bernabéu y García, 2017, p. 54).

Figura 5. **Modelo lógico de tabla de hechos para las notas de la Facultad de Ingeniería**



Fuente: elaboración propia, empleando Canva.

La tabla de hechos establecida para este *Data Warehouse* almacena hechos básicos y derivados. Los básicos son aquellos que ya están establecidos y las mismas fuentes de datos los proporcionan, tales como la zona, nota de laboratorio y examen final. Los derivados son los datos que se obtuvieron por medio de una operación, tales como nota final, mediana, desviación estándar, percentil 25, percentil 50 y percentil 75.

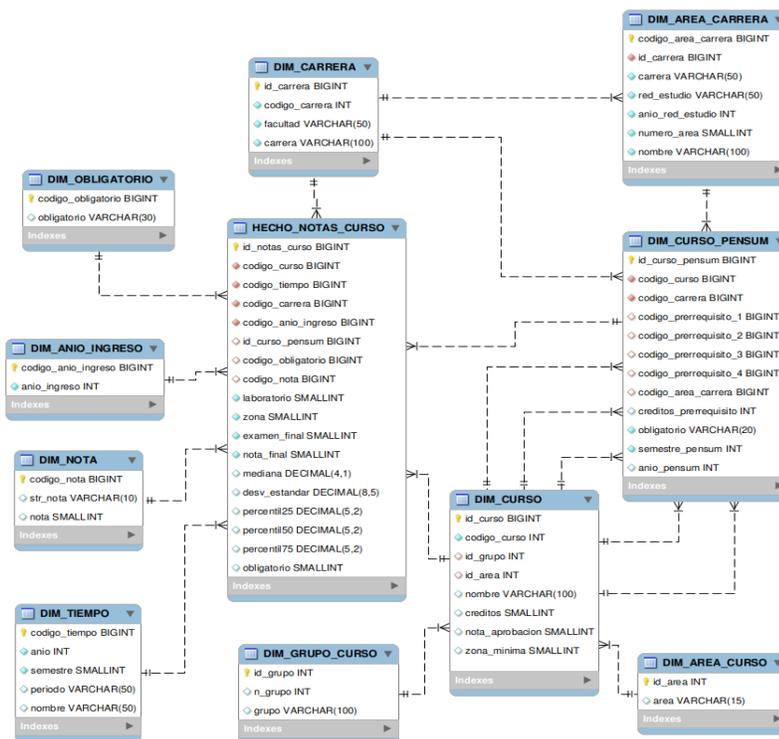
Almacenar estos datos calculados facilita las consultas, pero implica utilizar más espacio en el *Data Warehouse* y más tiempo de procesamiento en

el ETL, por lo que se debe almacenar únicamente los datos que se consideren convenientes para el análisis y obtener los demás indicadores al momento de realizar el cubo multidimensional.

- o Relaciones

Para finalizar con el modelo lógico se deben agregar las relaciones entre la tabla de hechos y las dimensiones.

Figura 6. **Modelo lógico de *Data Warehouse* para la Facultad de Ingeniería**



Fuente: elaboración propia, empleando MySQL Workbench.

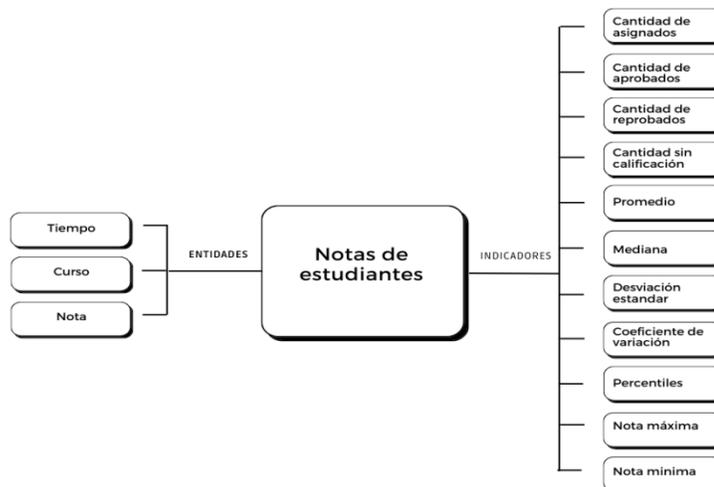
## 2.4.2.2. Diseño de *Data Warehouse* para el Programa Académico Preparatorio

Este diseño se describe a continuación:

### 2.4.2.2.1. Diseño de modelo conceptual para el Programa Académico Preparatorio

Para la creación del modelo conceptual se analizaron los datos proporcionados de las notas del Programa Académico Preparatorio ubicados en la tabla VII, donde se definieron las entidades a crear y los indicadores a calcular, tomando en cuenta la forma en la que se trabajó en la sección 2.4.2.1.2 para la Facultad de Ingeniería.

Figura 7. **Modelo conceptual del Programa Académico Preparatorio**



Fuente: elaboración propia, empleando Canva.

En el modelo del lado izquierdo se encuentran las entidades identificadas, las cuales son enlazadas por las notas de los estudiantes. Con las entidades y las notas obtenidas por los estudiantes se podrá calcular los indicadores situados del lado derecho.

#### **2.4.2.2.2. Diseño del modelo lógico para el Programa Académico Preparatorio**

Para la creación del modelo lógico se tomó en cuenta el procedimiento de la sección 2.4.2.1.2. del modelo lógico para la Facultad de Ingeniería y se realizaron los mismos pasos.

- Tablas de dimensiones

Son las siguientes:

- Dimensión tiempo

A partir de la entidad tiempo, en el modelo conceptual se definió el nombre de la dimensión, su clave principal, sus campos y la jerarquía, en este caso se definió la jerarquía año>semestre.

Figura 8. **Modelo lógico de la dimensión tiempo del Programa Académico Preparatorio**

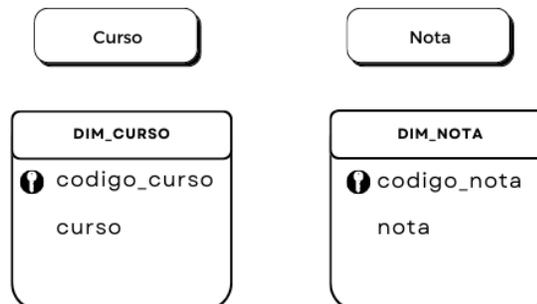


Fuente: elaboración propia, empleando Canva.

- Dimensiones independientes

A partir de las entidades en el modelo conceptual se definió para cada una el nombre de la dimensión, su clave principal, sus campos y sus jerarquías.

Figura 9. **Modelo lógico de las dimensiones del Programa Académico Preparatorio**



Fuente: elaboración propia, empleando Canva.

Al establecer las dimensiones del *Data Warehouse* se procede a definir el tipo de dato para cada campo y se establece según los datos recibidos que permitirán valores nulos.

Tabla XI. **Detalle de dimensiones en modelo lógico para el Programa Académico Preparatorio**

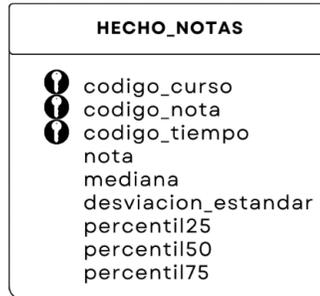
<b>Dimensión</b>	<b>Campo</b>	<b>Tipo</b>	<b>Permite nulos</b>
<b>Dim_tiempo</b>	Código_tiempo	Int	No
	Año	Int	No
	Semestre	Int	No
<b>Dim_curso</b>	Codigo_curso	Int	No
	Curso	Varchar(50)	No
<b>Dim_nota</b>	Codigo_nota	Int	No
	Nota	Int	No

Fuente: elaboración propia.

- Tabla de hechos

La tabla de hechos establecida para este *Data Warehouse* almacena hechos básicos y derivados. Como hecho básico se almacena la nota y como hechos derivados la mediana, desviación estándar, percentil 25, percentil 50 y percentil 75.

Figura 10. **Modelo lógico de la tabla de hechos para las notas del Programa Académico Preparatorio**

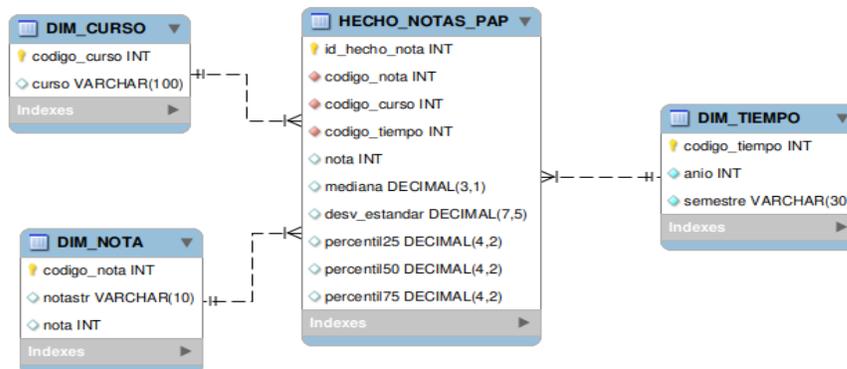


Fuente: elaboración propia, empleando Canva.

o Relaciones

Para finalizar con el modelo lógico se deben agregar las relaciones entre la tabla de hechos y las dimensiones.

Figura 11. **Modelo lógico del *Data Warehouse* para el Programa Académico Preparatorio**



Fuente: elaboración propia, empleando MySQL Workbench.

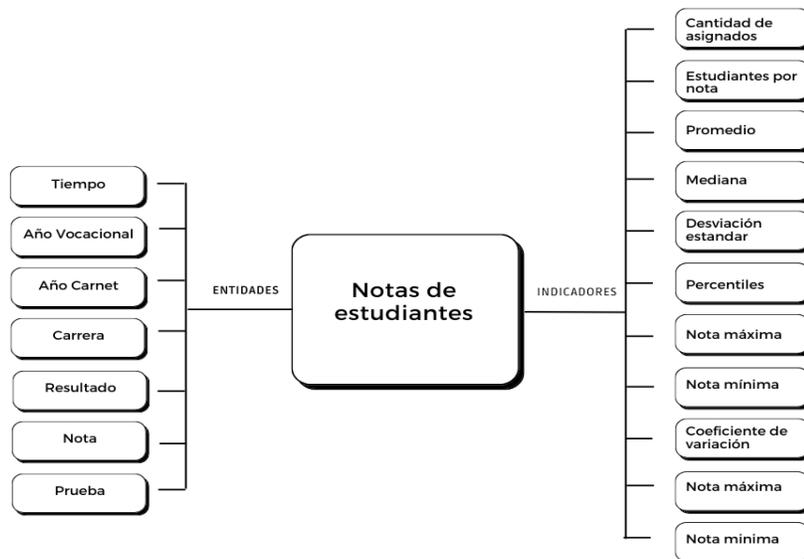
### 2.4.2.3. Diseño de *Data Warehouse* para las pruebas básicas

Este diseño se explica a continuación:

#### 2.4.2.3.1. Diseño del modelo conceptual para las pruebas básicas

Para la creación del modelo conceptual se analizaron los datos proporcionados de las notas satisfactorias de los estudiantes asignados a la Facultad de Ingeniería en las pruebas básicas de física y lenguaje, ubicados en la tabla VIII. Se definieron las entidades a crear y los indicadores a calcular, tomando en cuenta la forma en la que se trabajó en la sección 2.4.2.1.2 para la Facultad de Ingeniería.

Figura 12. Modelo conceptual de las pruebas básicas



Fuente: elaboración propia, empleando Canva.

En el modelo conceptual del lado izquierdo se encuentran las entidades definidas relacionadas a las notas de los estudiantes con las cuales se podrá realizar el cálculo de los indicadores situados al lado derecho.

#### **2.4.2.3.2.      Diseño del modelo lógico para las pruebas básicas**

Para la creación del modelo lógico se tomó en cuenta el procedimiento de la sección 2.4.2.1.2. del modelo lógico para la Facultad de Ingeniería y se realizaron los mismos pasos.

- Tablas de dimensiones

Son las siguientes:

- Dimensión tiempo

A partir de la entidad tiempo, en el modelo conceptual se definió el nombre de la dimensión, su clave principal y sus campos, en este caso se definió la jerarquía año>semestre>mes>día>fecha.

Figura 13. **Modelo lógico de la dimensión tiempo de las pruebas básicas**

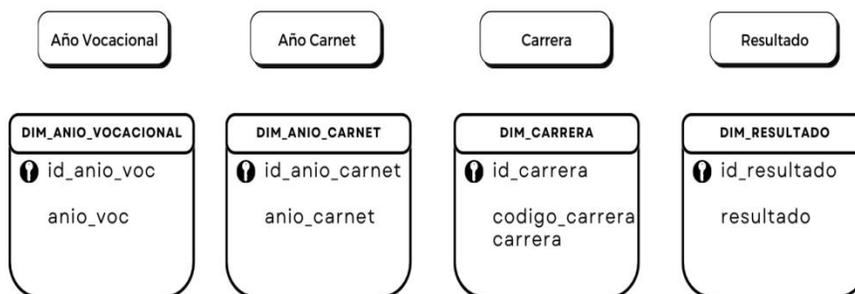


Fuente: elaboración propia, empleando Canva.

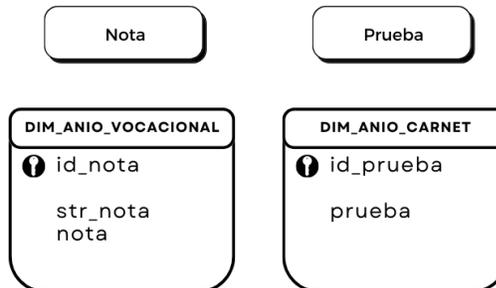
- Dimensiones independientes

A partir de las entidades en el modelo conceptual se definió para cada una el nombre de la dimensión, su clave principal, sus campos y sus jerarquías.

Figura 14. **Modelo lógico de las dimensiones de las pruebas básicas**



Continuación de la figura 14.



Fuente: elaboración propia, empleando Canva.

Al establecer las dimensiones del *Data Warehouse* se procede a definir el tipo de dato para cada campo y se establece según los datos recibidos que permitirán valores nulos.

Tabla XII. **Detalle de dimensiones en modelo lógico para las pruebas básicas**

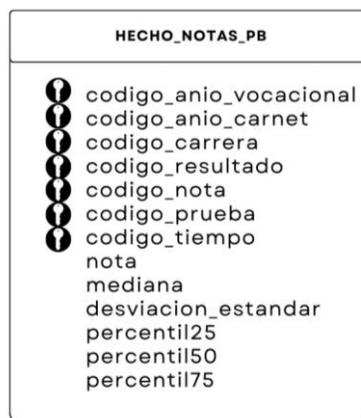
Dimensión	Campo	Tipo	Permite nulos
Dim_anio_vocacional	Id_anio_vocacional	Int	No
	anio_vocacional	Int	No
Dim_anio_carnet	id_anio_carnet	Int	No
	anio_carnet	Int	No
Dim_carrera	id_carrera	Int	No
	codigo_carrera	Int	No
	carrera	Varchar(100)	No
Dim_resultado	id_resultado	Int	No
	Resultado	Varchar(15)	No
Dim_nota	id_nota	Int	No
	str_nota	Varchar(10)	No
	Nota	Int	No
Dim_prueba	id_prueba	Int	No
	prueba	Varchar(30)	No

Fuente: elaboración propia.

- Tabla de hechos

Se presenta a continuación:

Figura 15. **Modelo lógico de la tabla de hechos para las notas de las pruebas básicas**



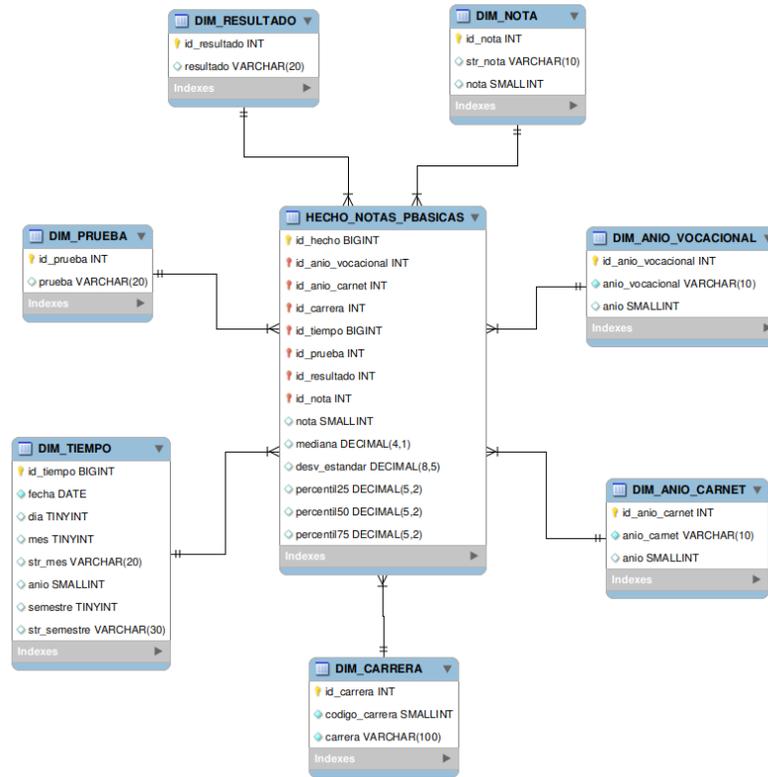
Fuente: elaboración propia, empleando Canva.

La tabla de hechos establecida para este *Data Warehouse* almacena hechos básicos y derivados. Como hecho básico se almacena la nota y como hechos derivados la mediana, desviación estándar, percentil 25, percentil 50 y percentil 75.

- Relaciones

Para finalizar con el modelo lógico se deben agregar las relaciones entre la tabla de hechos y las dimensiones.

Figura 16. **Modelo lógico de *Data Warehouse* para las notas de las pruebas básicas**



Fuente: elaboración propia, empleando MySQL Workbench.

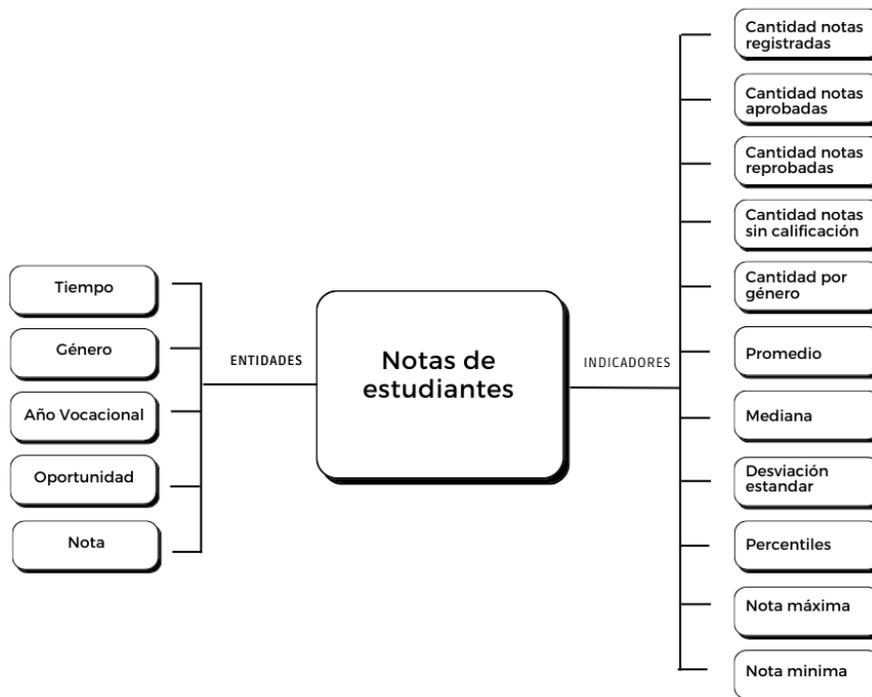
#### 2.4.2.4. **Diseño de *Data Warehouse* para la prueba específica**

Se explica en los siguientes incisos:

### 2.4.2.4.1. Diseño del modelo conceptual para la prueba específica

Para la creación del modelo conceptual se analizaron los datos proporcionados en la tabla IX de las notas de los aspirantes a ingresar a la Facultad de Ingeniería que realizaron la prueba específica de matemática. A partir de los datos se realizó un análisis para definir las entidades y los indicadores que se desea mostrar, tomando en cuenta la forma en la que se trabajó en la sección 2.4.2.1.2 para la Facultad de Ingeniería.

Figura 17. **Modelo conceptual de la prueba específica**



Fuente: elaboración propia, empleando Canva.

En el modelo conceptual del lado izquierdo se encuentran las entidades definidas relacionadas a las notas de los estudiantes con las cuales se podrá realizar el cálculo de los indicadores situados al lado derecho.

#### **2.4.2.4.2.      Diseño del modelo lógico para la prueba específica**

Para la creación del modelo lógico se tomó en cuenta el procedimiento de la sección 2.4.2.1.2. del modelo lógico para la Facultad de Ingeniería y se realizaron los mismos pasos.

- Tablas de dimensiones

Son las siguientes:

- Dimensión tiempo

A partir de la entidad tiempo, en el modelo conceptual se definió el nombre de la dimensión, su clave principal y su campo año, ya que es el único dato con respecto al tiempo proporcionado.

Figura 18. **Modelo lógico de la dimensión tiempo de la prueba específica**

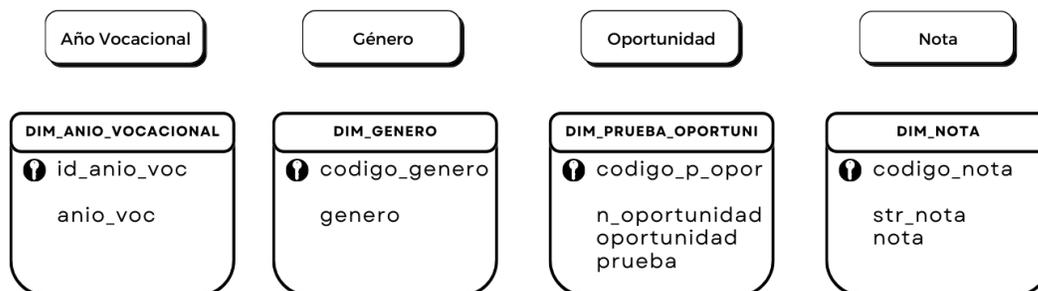


Fuente: elaboración propia, empleando Canva.

- Dimensiones independientes

A partir de las entidades en el modelo conceptual se definió para cada una el nombre de la dimensión, su clave principal, sus campos y sus jerarquías.

Figura 19. **Modelo lógico de las dimensiones de la prueba específica**



Fuente: elaboración propia, empleando Canva.

Al establecer las dimensiones del *Data Warehouse* se procede a definir el tipo de dato para cada campo y se establece según los datos recibidos que permitirán valores nulos.

Tabla XIII. **Detalle de dimensiones en modelo lógico para la prueba específica**

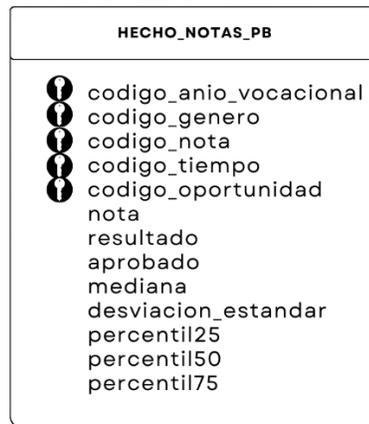
Dimensión	Campo	Tipo	Permite nulos
<b>Dim_anio_vocacional</b>	codigo_anio_vocacional	Int	No
	anio_vocacional	Int	No
<b>Dim_genero</b>	codigo_genero	Int	No
	genero	Varchar(20)	No
<b>Dim_prueba_oportunidad</b>	codigo_n_oportunidad	Int	No
	n_oportunidad	Varchar(50)	No
	oportunidad	Varchar(20)	Sí
	Prueba	Varchar(20)	Sí
<b>Dim_nota</b>	codigo_nota	Int	No
	Nota	Int	No

Fuente: elaboración propia.

- Tabla de hechos

La tabla de hechos establecida para este *Data Warehouse* almacena hechos básicos y derivados. Como hecho básico se almacena la nota, resultado y aprobado. Como hechos derivados la mediana, desviación estándar, percentil 25, percentil 50 y percentil 75.

Figura 20. **Modelo lógico de la tabla de hechos para las notas de la prueba específica**

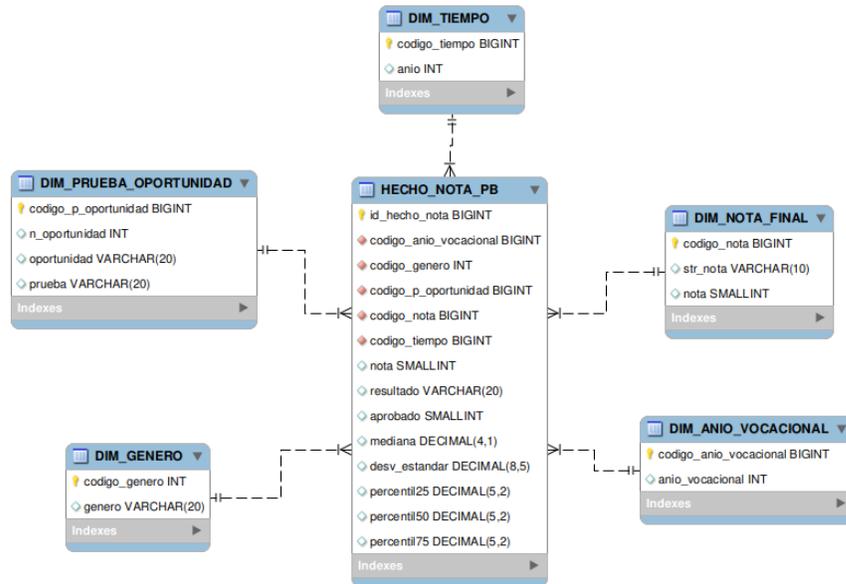


Fuente: elaboración propia, empleando Canva.

- Relaciones

Para finalizar con el modelo lógico se deben agregar las relaciones entre la tabla de hechos y las dimensiones.

Figura 21. **Modelo lógico del *Data Warehouse* para las notas de la prueba específica**



Fuente: elaboración propia, empleando MySQL Workbench.

### 2.4.3. Proceso ETL

Al tener construido el *Data Warehouse* para cada entidad, se procede a diseñar el proceso de ETL para cada uno, haciendo uso de la herramienta de código libre Pentaho. Pentaho es una plataforma que cuenta con la herramienta de Pentaho Data Integration (PDI), también conocida como Kettle. Esta herramienta permite implementar procesos de extracción, transformación y carga de datos (ETL), permitiendo a sus usuarios ahorrarse tiempo en realizar dichos procesos de manera manual (Cupas, 2022).

Esta herramienta permite crear dos tipos de archivos llamados transformaciones y *jobs*. Las transformaciones permiten crear el flujo del

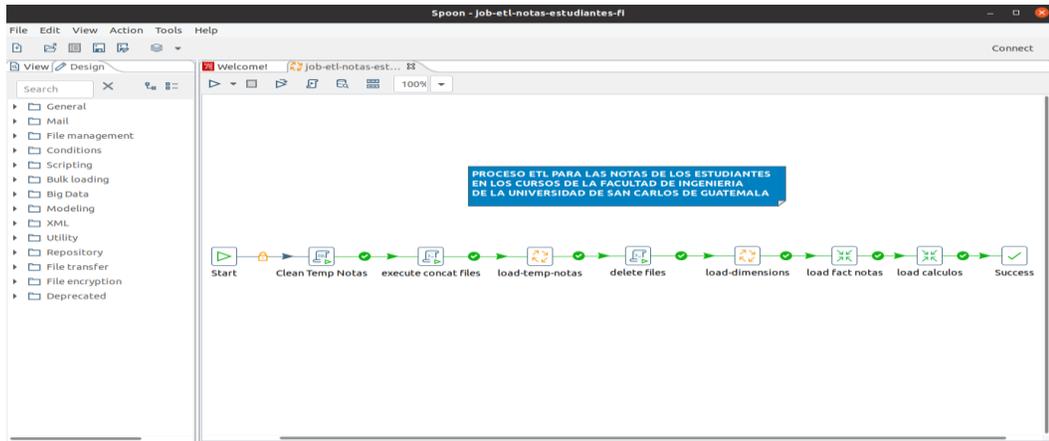
proceso de ETL, como al obtener los datos de una fuente de archivos, de una base de datos u otros medios, también se pueden transformar los datos y cargarlos en una ubicación específica. Los *jobs* permiten coordinar el flujo de todas las actividades de un proceso de ETL, indicando el orden en el que se debe ejecutar el proceso.

#### **2.4.3.1. Proceso de ETL para las notas de los estudiantes de la Facultad de Ingeniería**

El proceso de ETL para las notas de los estudiantes de la Facultad de Ingeniería está compuesto por transformaciones y *jobs* que permiten crear el flujo del programa. Para ejecutar este proceso se creó un *job* principal que marca el flujo del proceso y está compuesto por diferentes pasos en donde se extraerán las notas de los archivos Excel y se cargarán los datos seleccionados a una tabla temporal.

Desde la tabla temporal se cargarán los datos a las tablas de dimensiones, desde el nivel más general al más detallado, y se transformarán los datos en el proceso. Luego se cargarán los datos a la tabla de hecho y por último se calcularán los datos estadísticos con las notas previamente cargadas.

Figura 22. Proceso de ETL para la Facultad de Ingeniería



Fuente: elaboración propia, empleando Pentaho Data Integration.

Los componentes utilizados de la herramienta Pentaho Data Integration en el *job* principal se describen a continuación.

Tabla XIV. Componentes de *job* principal

Icono	Nombre	Función
	Start	Define donde iniciará la ejecución del <i>job</i> .
	SQL	Ejecuta comandos de sql en una determinada conexión a la base de datos.
	Shell	Permite ejecutar archivos externos en la <i>shell</i> .
	Job	Ejecuta un archivo <i>job</i> creado previamente.
	Transformation	Ejecuta un archivo de transformación creado previamente.
	Success	Define dónde termina la ejecución del <i>job</i> exitosamente.

Fuente: elaboración propia.

A continuación, se describen las tareas realizadas en el *job* principal para las notas de la Facultad de Ingeniería.

Tabla XV. **Tareas realizadas en el *job* principal para la Facultad de Ingeniería**

<b>Nombre tarea</b>	<b>Descripción</b>
<b>Start</b>	Inicia la ejecución de las tareas en el <i>job</i> .
<b>Clean Temp Notas</b>	Realiza una conexión a la base de datos del <i>Data Warehouse</i> de la Facultad de Ingeniería y limpia las tablas temporales.
<b>Execute concat files</b>	Ejecuta un archivo de Python en el que se verifica la estructura de los archivos, los procesa y mueve a otra carpeta para su análisis.
<b>Load Temp Notas</b>	Ejecuta un <i>job</i> creado que analiza los archivos de una carpeta definida y selecciona los datos a cargar en las tablas temporales.
<b>Delete files</b>	Ejecuta un archivo de Python que elimina los archivos que ya se cargaron a las tablas temporales.
<b>Load dimensions</b>	Ejecuta un <i>job</i> creado para la transformación y carga de datos que se encuentran en las tablas temporales a sus respectivas tablas de dimensiones de tiempo, año de ingreso de estudiantes, curso, área, grupo, nota y carrera.
<b>Load fact notas</b>	Ejecuta una transformación que cargará los datos a la tabla de hechos de las notas de los estudiantes.
<b>Load cálculos</b>	Ejecuta una transformación que realiza los cálculos estadísticos de la mediana, desviación estándar, percentiles 25, 50 y 75.
<b>Success</b>	Indica el fin de las tareas en el <i>job</i> de manera exitosa.

Fuente: elaboración propia.

### 2.4.3.2. Proceso de ETL para las notas de los estudiantes del curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio

El proceso de ETL para las notas de los estudiantes del curso de Matemática para Ingeniería del Programa Académico Preparatorio está compuesto, de igual manera que en la sección 2.4.3.1, por transformaciones y jobs. Se creó un *job* principal para marcar el flujo del proceso, el cual está compuesto por diferentes componentes que realizan las tareas indicadas.

Figura 23. Proceso de ETL para el Programa Académico Preparatorio



Fuente: elaboración propia, empleando Pentaho Data Integration.

Los componentes utilizados de la herramienta Pentaho Data Integration en el *job* principal para el Programa Académico Preparatorio son los mismos descritos en la tabla XII. A continuación se describen las tareas realizadas en el *job* principal para las notas del Programa Académico Preparatorio.

Tabla XVI. **Tareas realizadas en el *job* principal para el Programa Académico Preparatorio**

<b>Nombre tarea</b>	<b>Descripción</b>
<b>Start</b>	Inicia la ejecución de las tareas en el <i>job</i> .
<b>Delete temp</b>	Realiza una conexión a la base de datos del <i>Data Warehouse</i> del Programa Académico Preparatorio y limpia la tabla temporal.
<b>Shell</b>	Ejecuta un archivo de Python en el que se verifica la estructura de los archivos, los procesa y mueve a otra carpeta para su análisis.
<b>Load temp</b>	Ejecuta un <i>job</i> creado que analiza los archivos de una carpeta definida y selecciona los datos a cargar en la tabla temporal.
<b>Shell remove</b>	Ejecuta un archivo de Python que elimina los archivos que ya se cargaron a la tabla temporal.
<b>Load dimensiones</b>	Ejecuta un <i>job</i> creado para la transformación y carga de datos que se encuentran en la tabla temporal a sus respectivas tablas de dimensiones de tiempo, curso y nota.
<b>Load hecho</b>	Ejecuta una transformación que cargará los datos a la tabla de hechos de las notas de los estudiantes.
<b>Cálculos estadísticos</b>	Ejecuta una transformación que realiza los cálculos estadísticos de la mediana, desviación estándar, percentiles 25, 50 y 75.
<b>Success</b>	Indica el fin de las tareas en el <i>job</i> de manera exitosa.

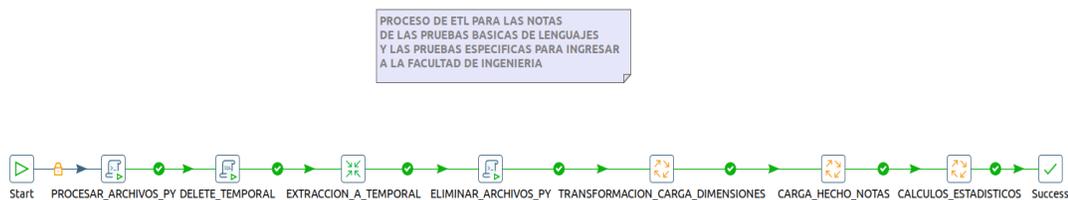
Fuente: elaboración propia.

### **2.4.3.3. Proceso de ETL para las notas de los estudiantes que aprobaron las pruebas básicas de física y lenguaje**

El proceso de ETL para las notas de los estudiantes que aprobaron las pruebas de conocimiento básico de física y lenguaje para ingresar a la Facultad de Ingeniería está compuesto, de igual manera que en la sección 2.4.3.1, por transformaciones y *jobs*. Se creó un *job* principal para marcar el flujo del

proceso, el cual está compuesto por diferentes componentes que realizan las tareas indicadas.

Figura 24. **Proceso de ETL para las pruebas básicas**



Fuente: elaboración propia, empleando Pentaho Data Integration.

Los componentes utilizados de la herramienta Pentaho Data Integration en el *job* principal para las pruebas básicas son los mismos descritos en la tabla XII. A continuación, se describen las tareas realizadas en el *job* principal para las notas de los aprobados en las pruebas básicas de física y lenguaje.

Tabla XVII. **Tareas realizadas en el *job* principal para las pruebas básicas de física y lenguaje**

Nombre tarea	Descripción
<b>Start</b>	Inicia la ejecución de las tareas en el <i>job</i> .
<b>Procesar archivos py</b>	Ejecuta un archivo de Python en el que se verifica la estructura de los archivos, los procesa y mueve a otra carpeta para su análisis.
<b>Delete temporal</b>	Realiza una conexión a la base de datos de las pruebas básicas y limpia la tabla temporal.
<b>Extracción a temporal</b>	Ejecuta una transformación creada que analiza los archivos Excel de una carpeta definida y selecciona los datos necesarios a cargar a la tabla temporal.
<b>Eliminar archivos py</b>	Ejecuta un archivo de Python que elimina los archivos Excel que ya se cargaron a la tabla temporal.

Continuación tabla XVII.

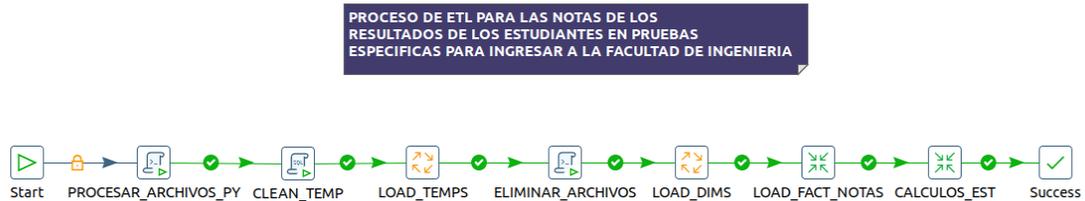
<b>Transformación dimensiones</b>	<b>carga</b>	<b>Ejecuta un <i>job</i> creado para la transformación y carga de datos que se encuentran en la tabla temporal a sus respectivas tablas de dimensiones de tiempo, año de número de carné, año de número de orientación vocacional, nota, prueba, carrera.</b>
<b>Carga hecho notas</b>		Ejecuta una transformación que cargará los datos a la tabla de hechos de las notas de los estudiantes.
<b>Cálculos estadísticos</b>		Ejecuta una transformación que realiza los cálculos estadísticos de la mediana, desviación estándar, percentiles 25, 50 y 75.
<b>Success</b>		Indica el fin de las tareas en el <i>job</i> de manera exitosa.

Fuente: elaboración propia.

#### **2.4.3.4. Proceso de ETL para las notas de los estudiantes que realizaron la prueba específica de matemática**

El proceso de ETL para las notas de los estudiantes que realizaron la prueba específica de matemática para ingresar a la Facultad de Ingeniería está compuesto, de igual manera que en la sección 2.4.3.1, por transformaciones y *jobs*. Se creó un *job* principal para marcar el flujo del proceso, el cual está compuesto por diferentes componentes que realizan las tareas indicadas.

Figura 25. **Proceso de ETL para la prueba específica de matemática**



Fuente: elaboración propia, empleando Pentaho Data Integration.

Los componentes utilizados de la herramienta Pentaho Data Integration en el *job* principal para la prueba específica son los mismos descritos en la tabla XII. A continuación, se describen las tareas realizadas en el *job* principal para las notas de quienes realizaron la prueba específica de matemática.

Tabla XVIII. **Tareas realizadas en el *job* principal para la prueba específica de matemática**

Nombre tarea	Descripción
<b>Start</b>	Inicia la ejecución de las tareas en el <i>job</i> .
<b>Procesar archivos py</b>	Ejecuta un archivo de Python en el que se verifica la estructura de los archivos, los procesa y mueve a otra carpeta para su análisis.
<b>Clean temp</b>	Realiza una conexión a la base de datos de la prueba específica y limpia la tabla temporal.
<b>Load temps</b>	Ejecuta un <i>job</i> creado que analiza los archivos Excel de una carpeta definida y selecciona los datos necesarios a cargar a la tabla temporal.
<b>Eliminar archivos</b>	Ejecuta un archivo de Python que elimina los archivos Excel que ya se cargaron a la tabla temporal.

Continuación tabla XVII.

<b>Load dims</b>	Ejecuta un <i>job</i> creado para la transformación y carga de datos que se encuentran en la tabla temporal a sus respectivas tablas de dimensiones de tiempo, año de número de orientación vocacional, nota, prueba y género.
<b>Load fact notas</b>	Ejecuta una transformación que carga los datos a la tabla de hechos de las notas de los estudiantes.
<b>Cálculos estadísticos</b>	Ejecuta una transformación que realiza los cálculos estadísticos de la mediana, desviación estándar, percentiles 25, 50 y 75.
<b>Success</b>	Indica el fin de las tareas en el <i>job</i> de manera exitosa.

Fuente: elaboración propia.

#### 2.4.4. Cubo multidimensional OLAP

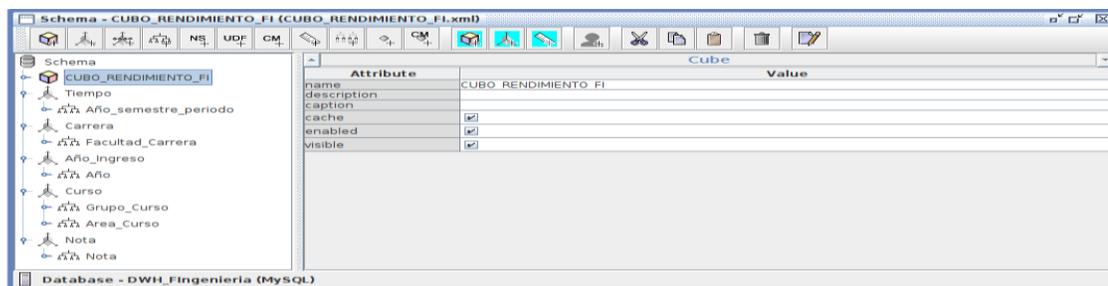
La información almacenada en los *Data Warehouses* será analizada por un cubo multidimensional OLAP (Online Analytic Processing) y se hará uso de la herramienta de código libre Pentaho Schema Workbench para la creación de la estructura del cubo. Esta herramienta permite la creación de cubos multidimensionales con el servidor OLAP Mondrian, el cual gestiona la comunicación con la base de datos, realiza búsquedas en menor tiempo y analiza grandes cantidades de datos.

Esta herramienta tiene la capacidad de crear cubos compuestos por dimensiones y métricas calculadas, los cuales se guardan como un archivo XML (Espinosa, 2010).

### 2.4.4.1. Cubo multidimensional para la Facultad de Ingeniería

Para la creación del cubo multidimensional de la Facultad de Ingeniería es necesario comprender el diseño del *Data Warehouse*. Como se puede observar en la figura 5, la tabla de hechos está conformada por 7 claves subrogadas que hacen referencia a las tablas de dimensiones, almacena los datos numéricos de las notas obtenidas por los estudiantes como la zona, zona en el laboratorio, examen final y nota final. También almacena la carrera y el año de ingreso del estudiante, el curso en el que obtuvo las notas, el pénsum al que pertenece el curso y si el curso era obligatorio para el estudiante o no.

Figura 26. Cubo multidimensional para la Facultad de Ingeniería



Fuente: elaboración propia, empleando Schema Workbench.

Los componentes utilizados de la herramienta Pentaho Schema Workbench en el cubo multidimensional se describen a continuación.

Tabla XIX. **Componentes del cubo multidimensional**

Icono	Nombre	Función
	Agregar cubo	Crea un cubo multidimensional.
	Agregar dimensión	Crea una dimensión que puede ser reutilizada en más de un cubo.
	Agregar uso de dimensión	Agrega una dimensión creada previamente para utilizarla en el cubo seleccionado.
	Agregar jerarquía	Crea una jerarquía entre las dimensiones o atributos.
	Agregar nivel	Crea un nivel en la jerarquía desde el nivel más alto al más bajo.
	Agregar métrica	Crea una métrica donde se realiza un cálculo sobre una columna de la tabla de hechos o sobre una expresión de sql.
	Agregar miembro calculado	Crea un miembro calculado donde se realiza un cálculo sobre los datos de otras métricas creadas previamente.

Fuente: elaboración propia.

Se debe crear las dimensiones que se desea que contenga el cubo multidimensional, estas hacen referencia a las tablas de dimensiones del *Data Warehouse* de la Facultad de Ingeniería. A cada dimensión se le debe especificar las jerarquías necesarias, las cuales contienen niveles para el análisis del modelo. Las dimensiones y jerarquías tienen importancia dentro del cubo multidimensional, ya que permiten aplicar filtros sobre las métricas y miembros calculados, obteniendo resultados más específicos.

Tabla XX. Dimensiones y jerarquías en cubo multidimensional para la Facultad de Ingeniería

Dimensión	Jerarquía	Niveles
<b>Tiempo</b>	Año>Semestre>Periodo	<ul style="list-style-type: none"> <li>• Año</li> <li>• Semestre</li> <li>• Periodo</li> </ul>
<b>Curso</b>	Grupo>Curso	<ul style="list-style-type: none"> <li>• Grupo</li> </ul>
	Área>Curso	<ul style="list-style-type: none"> <li>• Área</li> <li>• Curso</li> </ul>
<b>Carrera</b>	Facultad>Carrera	<ul style="list-style-type: none"> <li>• Facultad</li> <li>• Carrera</li> </ul>
<b>Año ingreso</b>	Año	<ul style="list-style-type: none"> <li>• Año</li> </ul>
<b>Nota</b>	Nota final	<ul style="list-style-type: none"> <li>• Nota</li> </ul>

Fuente: elaboración propia.

Los cubos tienen la capacidad de operar grandes cantidades de datos, los cuales serán mostrados a los interesados a través de métricas. Se debe realizar un análisis sobre los datos que se quiere mostrar a los interesados para hacer los cálculos. En este caso, como se desea medir el rendimiento de los estudiantes en los cursos de la Facultad de Ingeniería, se calcularon los siguientes datos:

Tabla XXI. Métricas en cubo multidimensional para la Facultad de Ingeniería

Nombre	Función de agregación	Descripción
<b>Asignados</b>	Count	Calcula el total de notas asignadas.
<b>Aprobados</b>	Sum	Calcula el total de notas aprobadas con valor arriba de 60.

Continuación tabla XXI.

<b>Cantidad reprobados</b>	<b>Sum</b>	<b>Calcula el total de notas reprobadas con valor debajo de 61.</b>
<b>Promedio</b>	Avg	Calcula el promedio de las notas finales.
<b>Promedio zona</b>	Avg	Calcula el promedio de las notas en la zona.
<b>Promedio zona laboratorio</b>	Avg	Calcula el promedio de las notas en la zona de laboratorio.
<b>Nota máxima final</b>	Max	Calcula la nota máxima obtenida en las notas finales.
<b>Nota mínima final</b>	Min	Calcula la nota mínima obtenida en las notas finales.
<b>Cantidad de cursos</b>	Count	Calcula el total de cursos registrados.
<b>Sin zona mínima</b>	Sum	Calcula el total de notas con valor de zona debajo de 36 y de nota de examen final con valor de 0.
<b>Desviación estándar</b>	Avg	Calcula el promedio de la desviación estándar calculada.
<b>Percentil 25</b>	Avg	Calcula el promedio del percentil 25 calculado.
<b>Percentil 50</b>	Avg	Calcula el promedio del percentil 50 calculado.
<b>Percentil 75</b>	Avg	Calcula el promedio del percentil 75 calculado.

Fuente: elaboración propia.

Tabla XXII. **Miembros calculados en cubo multidimensional para la Facultad de Ingeniería**

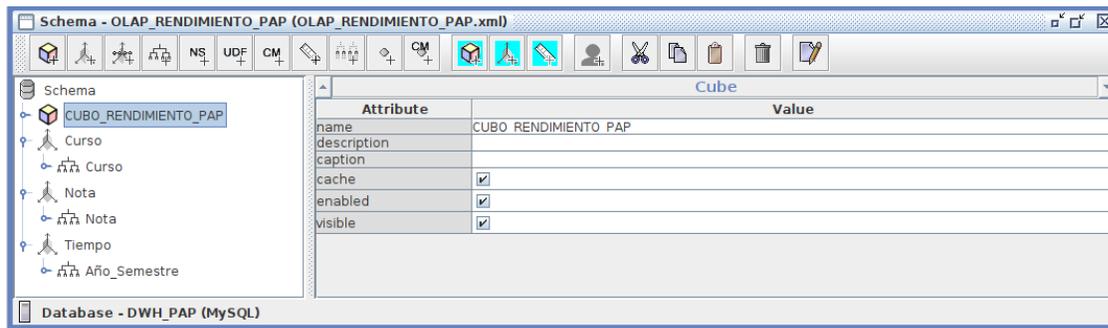
<b>Nombre</b>	<b>Descripción</b>
<b>Reprobados</b>	Calcula el total de reprobados diferenciándolos de quienes no obtuvieron zona mínima.
<b>Derecho a examen final</b>	Calcula el total de notas con derecho a examen final, diferenciando a los asignados de quienes no obtuvieron zona mínima.
<b>Porcentaje de aprobados</b>	Calcula el porcentaje del total de aprobados.
<b>Porcentaje de reprobados</b>	Calcula el porcentaje del total de reprobados que continuaron el curso.
<b>Porcentaje de sin zona mínima</b>	Calcula el porcentaje del total de notas que no obtuvieron zona mínima.
<b>Coeficiente de variación</b>	Calcula el coeficiente de variación de las notas de los estudiantes.
<b>Error de la media</b>	Calcula el error de la media sobre las notas válidas.

Fuente: elaboración propia.

#### **2.4.4.2. Cubo multidimensional para el Programa Académico Preparatorio**

Para la creación del cubo multidimensional del Programa Académico Preparatorio se tomó como base la tabla de hechos del *Data Warehouse* del mismo para seleccionar las dimensiones y métricas. Como se puede observar en la figura 10, la tabla de hechos está conformada por tres llaves subrogadas que hacen referencia a las tablas de dimensiones y almacena los datos numéricos de la nota y los cálculos estadísticos de la mediana, desviación estándar, percentil 25, 50 y 75.

Figura 27. **Cubo multidimensional para el Programa Académico Preparatorio**



Fuente: elaboración propia, empleando Schema Workbench.

Los componentes utilizados para la creación del cubo están descritos en la tabla XVII. A continuación se detallan las dimensiones y jerarquías agregadas en el cubo multidimensional:

Tabla XXIII. **Dimensiones y jerarquías en cubo multidimensional para el Programa Académico Preparatorio**

Dimensión	Jerarquía	Niveles
<b>Tiempo</b>	Año>Semestre	<ul style="list-style-type: none"> <li>• Año</li> <li>• Semestre</li> </ul>
<b>Nota</b>	Nota	<ul style="list-style-type: none"> <li>• Nota</li> </ul>
<b>Curso</b>	Curso	<ul style="list-style-type: none"> <li>• Curso</li> </ul>

Fuente: elaboración propia.

Para el cálculo de las métricas se realizó un análisis sobre los datos que aportaban valor a los interesados, para medir el rendimiento de los estudiantes del curso de Matemática Preparatoria para Ingeniería.

Tabla XXIV. **Métricas en cubo multidimensional para el Programa Académico Preparatorio**

<b>Nombre</b>	<b>Función de agregación</b>	<b>Descripción</b>
<b>Asignados</b>	Count	Calcula el total de asignados.
<b>Aprobados</b>	Sum	Calcula el total de notas aprobadas con valor arriba de 60.
<b>Promedio</b>	Avg	Calcula el promedio de las notas.
<b>Máximo</b>	Max	Calcula la nota máxima obtenida.
<b>Mínimo</b>	Min	Calcula la nota mínima obtenida.
<b>Cantidad de cursos</b>	Count	Calcula el total de cursos registrados.
<b>Sin calificación</b>	Sum	Calcula el total de notas sin calificación.
<b>Desviación estándar</b>	Avg	Calcula el promedio de la desviación estándar calculada.
<b>Percentil 25</b>	Avg	Calcula el promedio del percentil 25 calculado.
<b>Percentil 50</b>	Avg	Calcula el promedio del percentil 50 calculado.
<b>Percentil 75</b>	Avg	Calcula el promedio del percentil 75 calculado.

Fuente: elaboración propia.

Tabla XXV. **Miembros calculados en cubo multidimensional para el Programa Académico Preparatorio**

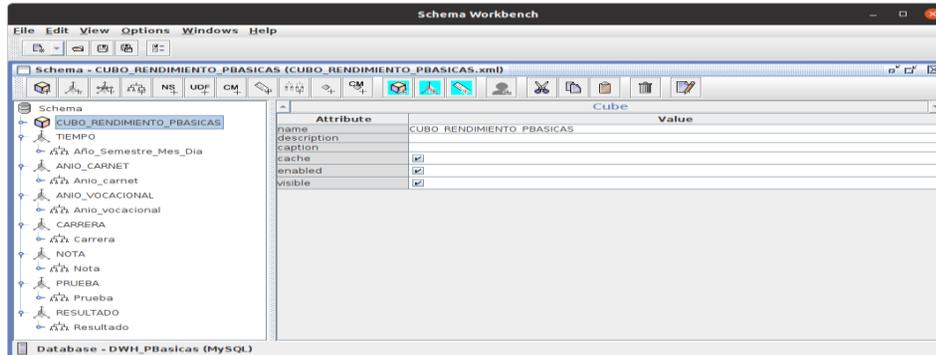
<b>Nombre</b>	<b>Descripción</b>
<b>Reprobados</b>	Calcula el total de reprobados diferenciándolos de quienes no tienen calificación.
<b>Porcentaje de aprobados</b>	Calcula el porcentaje del total de aprobados.
<b>Porcentaje de reprobados</b>	Calcula el porcentaje del total de reprobados que continuaron el curso.
<b>Porcentaje de sin calificación</b>	Calcula el porcentaje del total de notas del que no se registró calificación.
<b>Coefficiente de variación</b>	Calcula el coeficiente de variación de las notas de los estudiantes.
<b>Error de la media</b>	Calcula el error de la media sobre las notas válidas.

Fuente: elaboración propia.

#### **2.4.4.3. Cubo multidimensional para las pruebas básicas**

Para la creación del cubo multidimensional para las pruebas básicas se tomó como base la tabla de hechos del *Data Warehouse* del mismo para seleccionar las dimensiones y métricas. Como se puede observar en la figura 15, la tabla de hechos está conformada por siete llaves subrogadas que hacen referencia a las tablas de dimensiones y almacena los datos numéricos de las notas de las pruebas básicas de física y lenguaje y los cálculos estadísticos de la mediana, desviación estándar, percentil 25, 50 y 75.

Figura 28. **Cubo multidimensional para las pruebas básicas**



Fuente: elaboración propia, empleando Schema Workbench.

Los componentes utilizados para la creación del cubo están descritos en la tabla XVII. A continuación se detallan las dimensiones y jerarquías agregadas en el cubo multidimensional:

Tabla XXVI. **Dimensiones y jerarquías en cubo multidimensional para las pruebas básicas**

Dimensión	Jerarquía	Niveles
<b>Tiempo</b>	Año>Semestre>Mes>Día	<ul style="list-style-type: none"> <li>• Año</li> <li>• Semestre</li> <li>• Mes</li> <li>• Día</li> </ul>
<b>Año de ingreso</b>	Año	<ul style="list-style-type: none"> <li>• Año</li> </ul>
<b>Año de número de orientación vocacional</b>	Año	<ul style="list-style-type: none"> <li>• Año</li> </ul>
<b>Carrera</b>	Carrera	<ul style="list-style-type: none"> <li>• Carrera</li> </ul>
<b>Nota</b>	Nota	<ul style="list-style-type: none"> <li>• Nota</li> </ul>
<b>Prueba</b>	Prueba	<ul style="list-style-type: none"> <li>• Prueba</li> </ul>
<b>Resultado</b>	Resultado	<ul style="list-style-type: none"> <li>• Resultado</li> </ul>

Fuente: elaboración propia.

Para el cálculo de las métricas se realizó un análisis sobre los datos que aportaban valor a los interesados, para medir el rendimiento de los estudiantes de la Facultad de Ingeniería que aprobaron las pruebas básicas de física y lenguaje.

Tabla XXVII. **Métricas en cubo multidimensional para las pruebas básicas**

<b>Nombre</b>	<b>Función de agregación</b>	<b>Descripción</b>
<b>Asignados</b>	Count	Calcula el total de asignados.
<b>Promedio</b>	Avg	Calcula el promedio de las notas.
<b>Máximo</b>	Max	Calcula la nota máxima obtenida.
<b>Mínimo</b>	Min	Calcula la nota mínima obtenida.
<b>Mediana</b>	Avg	Calcula la mediana de las notas.
<b>Desviación estándar</b>	Avg	Calcula el promedio de la desviación estándar calculada.
<b>Percentil 25</b>	Avg	Calcula el promedio del percentil 25 calculado.
<b>Percentil 50</b>	Avg	Calcula el promedio del percentil 50 calculado.
<b>Percentil 75</b>	Avg	Calcula el promedio del percentil 75 calculado.

Fuente: elaboración propia.

Tabla XXVIII. **Miembros calculados en cubo multidimensional para las pruebas básicas**

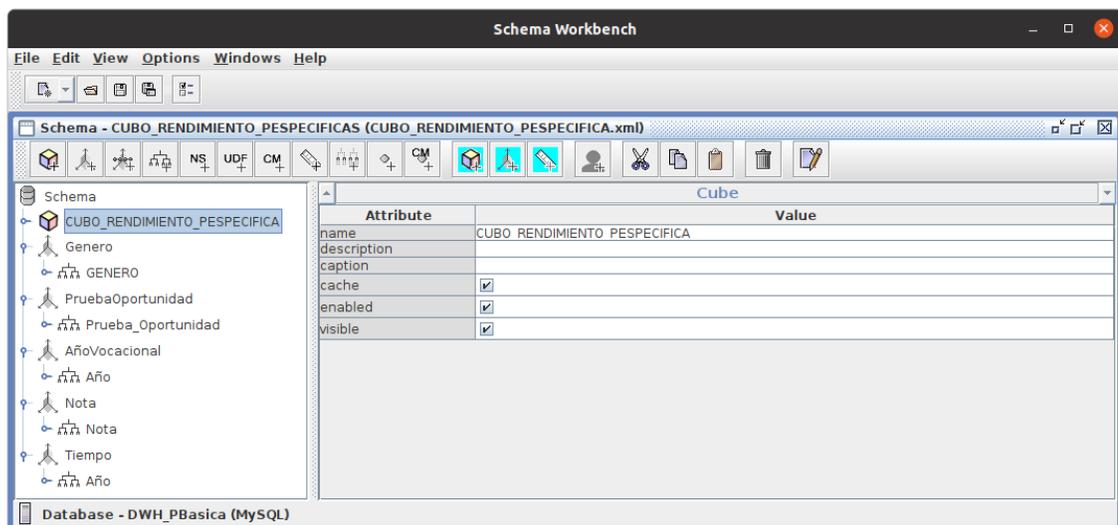
<b>Nombre</b>	<b>Descripción</b>
<b>Coefficiente de variación</b>	Calcula el coeficiente de variación de las notas de los estudiantes.

Fuente: elaboración propia.

#### 2.4.4.4. Cubo multidimensional para la prueba específica

Para la creación del cubo multidimensional para la prueba específica de matemática se tomó como base la tabla de hechos del *Data Warehouse* del mismo, para seleccionar las dimensiones y métricas. Como se puede observar en la figura 20, la tabla de hechos está conformada por cinco llaves subrogadas que hacen referencia a las tablas de dimensiones y almacena los datos numéricos de las notas de la prueba específica de matemática y los cálculos estadísticos de la mediana, desviación estándar, percentil 25, 50 y 75.

Figura 29. Cubo multidimensional para las pruebas específicas



Fuente: elaboración propia, empleando Schema Workbench.

Los componentes utilizados para la creación del cubo están descritos en la tabla XVII. A continuación se detallan las dimensiones y jerarquías agregadas en el cubo multidimensional.

Tabla XXIX. Dimensiones y jerarquías en cubo multidimensional para la prueba específica

Dimensión	Jerarquía	Niveles
<b>Género</b>	Género	• Género
<b>Prueba</b>	Prueba>Oportunidad	• Prueba • Oportunidad
<b>Año de número de orientación vocacional</b>	Año	• Año
<b>Nota</b>	Nota	• Nota
<b>Tiempo</b>	Año	• Año

Fuente: elaboración propia.

Para el cálculo de las métricas se realizó un análisis sobre los datos que aportaban valor a los interesados, para medir el rendimiento de los estudiantes que realizaron la prueba específica de matemática.

Tabla XXX. Métricas en cubo multidimensional para la prueba específica

Nombre	Función de agregación	Descripción
<b>Estudiantes</b>	Count	Calcula el total de estudiantes que realizaron la prueba.
<b>Aprobados</b>	Count	Calcula el total de estudiantes que obtuvieron resultado satisfactorio.
<b>Cantidad de reprobados</b>	Count	Calcula el total de estudiantes que obtuvieron resultado insatisfactorio.
<b>Sin calificación</b>	Count	Calcula el total de estudiantes del que no se agregó su calificación.
<b>Mujeres</b>	Count	Calcula el total de mujeres.
<b>Hombres</b>	Count	Calcula el total de hombres.
<b>Promedio</b>	Avg	Calcula el promedio de las notas.

Continuación tabla XXX.

<b>Máximo</b>	<b>Max</b>	<b>Calcula la nota máxima obtenida.</b>
<b>Mínimo</b>	Min	Calcula la nota mínima obtenida.
<b>Mediana</b>	Avg	Calcula la mediana de las notas.
<b>Desviación estándar</b>	Avg	Calcula el promedio de la desviación estándar calculada.
<b>Percentil 25</b>	Avg	Calcula el promedio del percentil 25 calculado.
<b>Percentil 50</b>	Avg	Calcula el promedio del percentil 50 calculado.
<b>Percentil 75</b>	Avg	Calcula el promedio del percentil 75 calculado.

Fuente: elaboración propia.

Tabla XXXI. **Miembros calculados en cubo multidimensional para las pruebas básicas**

<b>Nombre</b>	<b>Descripción</b>
<b>Reprobados</b>	Calcula el total de reprobados diferenciando a quienes no se les agregó su calificación.
<b>Porcentaje de aprobados</b>	Calcula el porcentaje de aprobados.
<b>Porcentaje de reprobados</b>	Calcula el porcentaje de reprobados.
<b>Porcentaje de sin calificación</b>	Calcula el porcentaje de sin calificación.
<b>Coefficiente de variación</b>	Calcula el coeficiente de variación de las notas de los estudiantes.

Fuente: elaboración propia.

#### 2.4.5. Cuadros de mando

Se utilizó la herramienta de código libre Pentaho Business Analytics, también conocida como Pentaho Server, y se utilizó dentro de la herramienta el

*plugin* Visualizer, desarrollado por Datafor, ya que permite la creación de cuadros de mando interactivos de manera intuitiva. Al utilizar Visualizer para la creación de cuadros de mando y Pentaho Schema Workbench para la creación de cubos multidimensionales, se facilitó su implementación, puesto que ambos utilizan Mondrian como motor multidimensional (Bernabéu, 2022).

Los cuadros de mando implementados tienen las siguientes características:

- Proporcionan capacidades de filtrado para limitar los datos y hacerlos más específicos en los reportes.
- Proporcionan la capacidad de exportar los datos de los reportes en formato Excel, CSV o como imagen.
- Algunos reportes tienen la capacidad de realizar un análisis sobre sus datos calculando la media y mediana.
- Algunos reportes permiten visualizar sus datos por medio de tablas.

#### **2.4.5.1. Cuadros de mando para la Facultad de Ingeniería**

Para la creación del cuadro de mando para la Facultad de Ingeniería se diseñó la presentación y se definieron los reportes que son relevantes para mostrar a los interesados. Estos reportes se basan en las notas de los estudiantes de la Facultad de Ingeniería, según el año y periodo en que se asignó el curso. Las métricas mostradas se obtuvieron del cubo multidimensional para la Facultad de Ingeniería en la sección 2.4.4.1.

Figura 30. Cuadro de mando para el monitoreo del rendimiento estudiantil de la Facultad de Ingeniería



Fuente: elaboración propia, empleando Pentaho Business Analytics

El cuadro de mando está conformado por dos pestañas para visualizar los reportes del rendimiento estudiantil para la Facultad de Ingeniería, una pestaña contiene los reportes generales que muestran datos de todos los cursos y la otra pestaña contiene los reportes por curso que muestran datos específicos para un curso seleccionado en un determinado año y periodo. Cuenta con la opción de aplicar filtros para especificar cuáles datos se prefiere visualizar y contiene tablas con datos estadísticos para medir el rendimiento estudiantil.

A continuación, se describen los reportes generales en la tabla, así como los reportes por curso:

Tabla XXXII. **Reportes generales para el monitoreo del rendimiento estudiantil de la Facultad de Ingeniería**

<b>Nombre</b>	<b>Descripción</b>
<b>Cantidad de notas registradas, aprobadas, reprobadas y sin zona mínima por carrera</b>	Gráfica de barras donde se puede visualizar las notas aprobadas, reprobadas y sin zona mínima por carrera.
<b>Porcentaje de aprobados, reprobados y sin zona mínima</b>	Gráfico circular que muestra el porcentaje de las notas aprobadas, las notas reprobadas y las notas sin zona mínima registradas en los cursos de la Facultad de Ingeniería.
<b>Aprobados por carrera</b>	Gráfico circular que muestra el porcentaje de notas aprobadas por carrera en los cursos de la Facultad de Ingeniería.
<b>Promedios por carrera</b>	Gráfico de área que muestra el promedio de la nota final, de zona, de laboratorio y de examen final por carrera de las notas en los cursos de la Facultad de Ingeniería.
<b>Promedio por año de ingreso</b>	Gráfica de barras que muestra el promedio de las notas finales por el año de ingreso de los estudiantes en los cursos de la Facultad de Ingeniería.
<b>Aprobados por año de ingreso</b>	Gráfico circular que muestra el porcentaje de las notas aprobadas por año de ingreso en los cursos de la Facultad de Ingeniería.
<b>Promedio de notas por grupo</b>	Gráfica de barras que muestra el promedio de las notas finales por grupo de los cursos de la Facultad de Ingeniería.
<b>Promedio de notas por área y curso</b>	Gráfica de barras que muestra el promedio de las notas finales por grupo de los cursos de la Facultad de Ingeniería.
<b>Estadísticos por año de ingreso</b>	Tabla que muestra los resultados de análisis estadísticos por año de ingreso en los cursos de la Facultad de Ingeniería.
<b>Estadísticos por curso</b>	Tabla que muestra los resultados de análisis estadísticos por los cursos de la Facultad de Ingeniería.

Fuente: elaboración propia.

Tabla XXXIII. **Reportes por curso para el monitoreo del rendimiento estudiantil de la Facultad de Ingeniería**

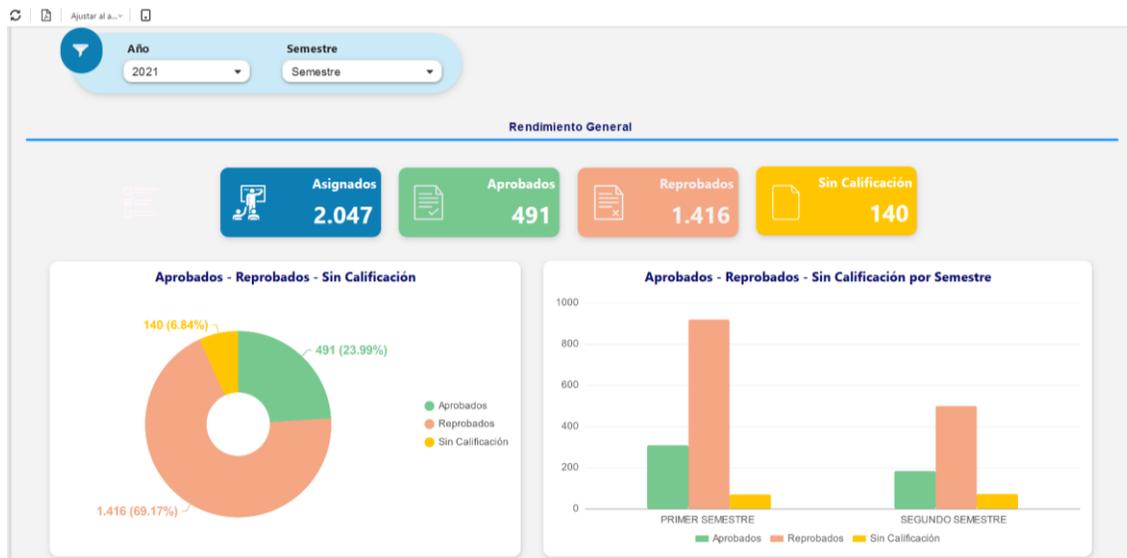
<b>Nombre</b>	<b>Descripción</b>
<b>Estadísticos del curso</b>	Tabla que muestra los resultados de análisis estadísticos en el curso seleccionado.
<b>Promedios por carrera</b>	Gráfico de barras que muestra el promedio de nota final junto con una gráfica de líneas que muestra el promedio de zona, de examen final y de laboratorio que obtuvieron los estudiantes en el curso seleccionado.
<b>Porcentaje de aprobados, reprobados y sin zona mínima</b>	Gráfico circular que muestra el porcentaje de las notas aprobadas, las notas reprobadas y las notas sin zona mínima registradas en el curso seleccionado.
<b>Cantidad de estudiantes por notas</b>	Gráfica de área que muestra la cantidad de estudiantes por nota que obtuvieron en el curso seleccionado.
<b>Promedio por año de ingreso</b>	Gráfica de barras que muestra el promedio de nota final que obtuvieron los estudiantes por año de ingreso en el curso seleccionado.
<b>Top de mejores notas por año de ingreso</b>	Gráfica de barras que muestra las notas máximas en las notas finales, notas de zona y de examen final que obtuvieron los estudiantes por año de ingreso en el curso.
<b>Promedio de laboratorio</b>	Gráfica de velocímetro que muestra como indicador el promedio de notas en el laboratorio del curso.
<b>Promedio de zona</b>	Gráfica de velocímetro que muestra como indicador el promedio de las zonas del curso.
<b>Promedio de examen final</b>	Gráfica de velocímetro que muestra como indicador el promedio de las notas del examen final del curso.
<b>Estadísticos por carrera y año de ingreso de los estudiantes</b>	Tabla que muestra los resultados de análisis estadísticos por carrera y año de ingreso de los estudiantes en el curso seleccionado.

Fuente: elaboración propia.

### 2.4.5.2. Cuadros de mando para el Programa Académico Preparatorio

Para la creación del cuadro de mando para el Programa Académico Preparatorio se diseñó la presentación y se definieron los reportes que son relevantes para mostrar a los interesados. Estos reportes se basan en las notas de los estudiantes del curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio, según el año y semestre en que se asignó el curso. Las métricas mostradas se obtuvieron del cubo multidimensional en la sección 2.4.4.2.

Figura 31. Cuadro de mando para el monitoreo del rendimiento estudiantil del curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio



Fuente: elaboración propia, empleando Pentaho Business Analytic.

El cuadro de mando está conformado por una pestaña para visualizar los reportes del rendimiento estudiantil en el curso de Matemática Preparatoria para Ingeniería, donde se muestra de manera general el rendimiento de los estudiantes filtrado por el año y semestre de las notas. Muestra los indicadores de total de asignados, aprobados, reprobados y sin calificación, así como reportes sobre los indicadores y tablas con datos estadísticos sobre las notas.

A continuación, se describen los reportes del curso:

**Tabla XXXIV. Reportes para el monitoreo del rendimiento estudiantil del curso de Matemática para Ingeniería del Programa Académico Preparatorio**

<b>Nombre</b>	<b>Descripción</b>
<b>Porcentaje de aprobados, reprobados y sin calificación</b>	Gráfico circular que muestra el total y el porcentaje de las notas aprobadas, reprobadas y sin calificación registradas.
<b>Cantidad de aprobados, reprobados y sin zona mínima</b>	Gráfica de barras que indica la cantidad de notas aprobadas, reprobadas y sin calificación registradas por semestre.
<b>Cantidad de estudiantes por notas</b>	Gráfica de área que muestra la cantidad de estudiantes por nota registrada.
<b>Estadísticos</b>	Tabla que muestra los resultados de análisis estadísticos por semestre.

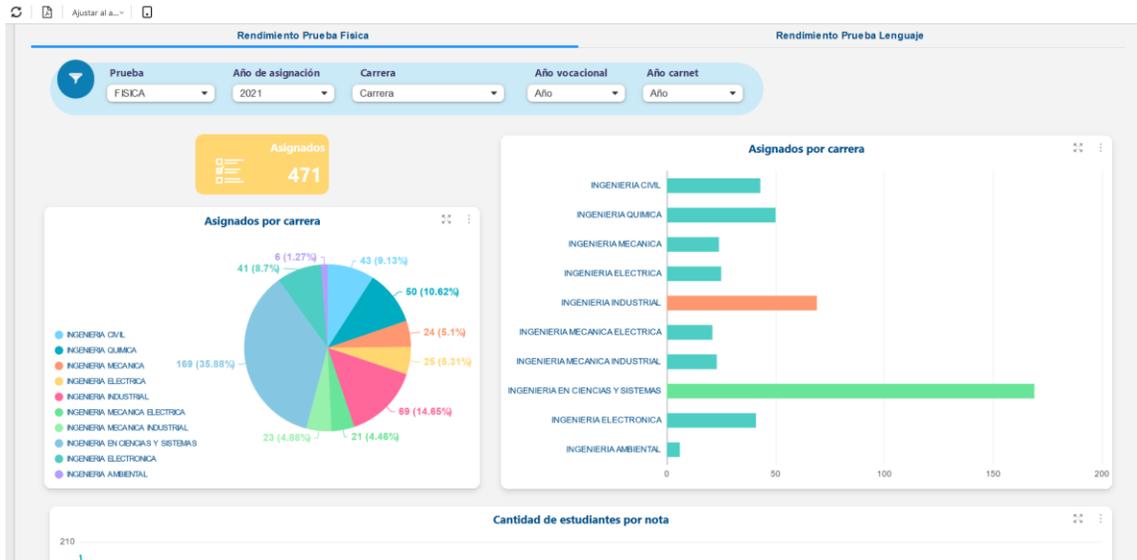
Fuente: elaboración propia.

### **2.4.5.3. Cuadros de mando para las pruebas básicas**

Para la creación del cuadro de mando de las pruebas básicas se diseñó la presentación y se definieron los reportes sobre las notas de los estudiantes que aprobaron las pruebas básicas de física y lenguaje, según su fecha de asignación a la Facultad de Ingeniería. Las métricas utilizadas en los reportes se obtuvieron por medio del cubo multidimensional realizado en la sección 2.4.4.3.

Por la confidencialidad de los datos de los estudiantes únicamente se muestra los primeros cuatro dígitos del registro académico, es decir, el año de ingreso de los estudiantes. Por lo tanto, los reportes son de manera general para monitorear el rendimiento estudiantil, según su año de ingreso y notas obtenidas. A continuación, se muestra un fragmento del cuadro de mando para las pruebas básicas.

Figura 32. Cuadro de mando para el monitoreo del rendimiento estudiantil para las pruebas básicas de física y lenguaje



Fuente: elaboración propia, empleando Pentaho Business Analytics

El cuadro de mando está conformado por dos pestañas, las cuales separan los reportes de las pruebas básicas de física y lenguaje, donde se permite filtrar por el año de asignación de los estudiantes a la Facultad de Ingeniería, la carrera del estudiante, el año del número de orientación vocacional y el año del registro académico. Los interesados pueden visualizar el indicador del total de asignados, reportes sobre el rendimiento estudiantil y tablas con datos estadísticos por fecha de asignación.

Tabla XXXV. **Reportes para el monitoreo del rendimiento estudiantil de los estudiantes aprobados de las pruebas básicas de física y lenguaje**

<b>Nombre</b>	<b>Descripción</b>
<b>Porcentaje de asignados por carrera</b>	Gráfico circular que muestra el porcentaje de los estudiantes por carrera que fueron asignados a la Facultad de Ingeniería que aprobaron las pruebas básicas.
<b>Cantidad de asignados por carrera</b>	Gráfico de barras que muestra el total de los estudiantes por carrera que fueron asignados a la Facultad de Ingeniería que aprobaron las pruebas básicas.
<b>Cantidad de estudiantes por notas</b>	Gráfica de área que muestra la cantidad de los estudiantes por nota que obtuvieron en las pruebas básicas.
<b>Promedio por año del número de orientación vocacional</b>	Gráfica de barras que muestra el promedio de nota que obtuvieron los estudiantes en las pruebas básicas por año del número de orientación vocacional.
<b>Promedio por año de asignación</b>	Gráfica de barras que muestra el promedio de las notas que obtuvieron los estudiantes en las pruebas básicas por año de asignación a la Facultad de Ingeniería.
<b>Estadísticos por año del número de orientación vocacional</b>	Tabla que muestra los resultados de los análisis estadísticos por el año del número de orientación vocacional de los asignados a la Facultad de Ingeniería en las pruebas básicas.
<b>Estadísticos por año de ingreso</b>	Tabla que muestra los resultados de análisis estadísticos en las pruebas básicas por el año de asignación a la Facultad de Ingeniería.
<b>Estadísticos por asignación de Escuela de Ciencias</b>	Tabla que indica los resultados de análisis estadísticos en las pruebas básicas por la fecha de asignación a la Facultad de Ingeniería en la Escuela de Ciencias.

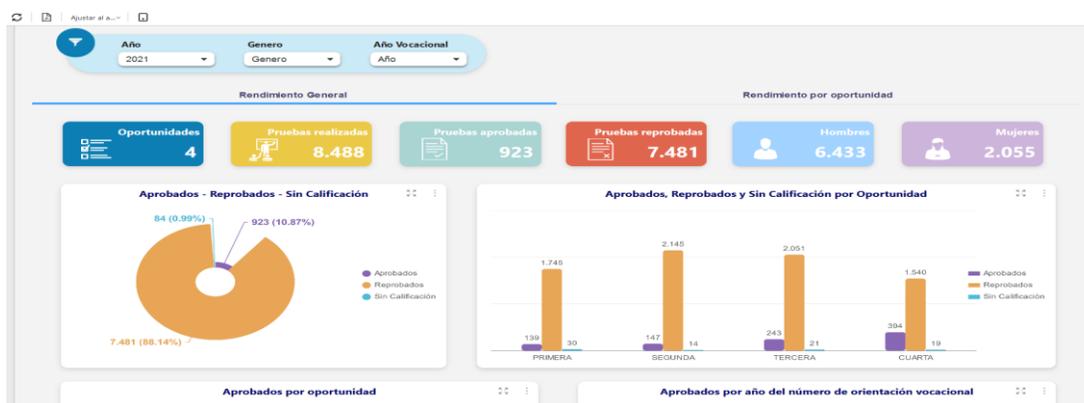
Fuente: elaboración propia.

#### 2.4.5.4. Cuadros de mando para la prueba específica

Para la creación del cuadro de mando de la prueba específica se diseñó la presentación y se definieron los reportes sobre las notas de los estudiantes que realizaron la prueba específica de matemática según el año y oportunidad. Las métricas utilizadas en los reportes se obtuvieron por medio del cubo multidimensional realizado en la sección 2.4.4.4.

Por la confidencialidad de los datos de los aspirantes a ingresar a la universidad, únicamente se muestra los primeros cuatro dígitos del número de orientación vocacional, es decir, el año en el cual inició el proceso de ingreso. Por lo tanto, los reportes son de manera general para monitorear el rendimiento de los aspirantes, según el año de inicio del proceso de ingreso y las notas obtenidas. A continuación, se muestra un fragmento del cuadro de mando para la prueba específica de matemática:

Figura 33. Cuadro de mando para el monitoreo del rendimiento estudiantil para la prueba específica de matemática



Fuente: elaboración propia, empleando Pentaho Business Analytics

El cuadro de mando está conformado por dos pestañas, las cuales separan los reportes generales y los reportes por oportunidad en la prueba específica de matemática, donde se permite filtrar por año en que se realizó la prueba, el género de los estudiantes y el año del número de orientación vocacional. La pestaña del rendimiento general muestra los datos de todas las oportunidades registradas para el año filtrado, donde los interesados pueden visualizar los indicadores, los totales de oportunidades, pruebas realizadas, pruebas aprobadas, pruebas reprobadas y la cantidad de hombres y mujeres.

Tabla XXXVI. **Reportes generales para el monitoreo del rendimiento estudiantil de los estudiantes que realizaron la prueba específica de matemática**

<b>Nombre</b>	<b>Descripción</b>
<b>Porcentaje de aprobados, reprobados y sin calificación</b>	Gráfico circular que muestra la cantidad y el porcentaje de los estudiantes que aprobaron, reprobaron y que no tuvieron calificación para la prueba específica en las todas las oportunidades.
<b>Cantidad de aprobados, reprobados y sin calificación por oportunidad</b>	Gráfico de barras que muestra el total de estudiantes que aprobaron, reprobaron y que no tuvieron calificación para la prueba específica en todas las oportunidades.
<b>Porcentaje de aprobados por oportunidad</b>	Gráfico circular que muestra la cantidad y el porcentaje de los estudiantes que aprobaron la prueba específica por oportunidad.
<b>Porcentaje de aprobados por año del número de orientación vocacional</b>	Gráfico circular que muestra la cantidad y el porcentaje de los estudiantes que aprobaron la prueba específica por el año del número de orientación vocacional.
<b>Cantidad de estudiantes por notas</b>	Gráfica de área que muestra la cantidad de estudiantes por notas que obtuvieron en la prueba específica de matemática.

Continuación tabla XXXVI.

<b>Promedio de notas por año del número de orientación vocacional</b>	<b>Gráfica de barras que muestra el promedio de nota que obtuvieron los estudiantes por año del número de orientación vocacional.</b>
<b>Estadísticos por año del número de orientación vocacional</b>	Tabla que muestra los resultados de análisis estadísticos por el año del número de orientación vocacional.
<b>Estadísticos</b>	Tabla que muestra los resultados de análisis estadísticos por oportunidades de la prueba específica de matemática.

Fuente: elaboración propia.

**Tabla XXXVII. Reportes por oportunidad para el monitoreo del rendimiento estudiantil de los estudiantes que realizaron la prueba específica de matemática**

<b>Nombre</b>	<b>Descripción</b>
<b>Estadísticos</b>	Tabla que muestra los resultados de análisis estadísticos por oportunidad de la prueba específica de matemática.
<b>Porcentaje de aprobados, reprobados y sin calificación</b>	Gráfico circular que muestra la cantidad y el porcentaje de los estudiantes que aprobaron, reprobaron y no obtuvieron calificación en la oportunidad filtrada de la prueba específica de matemática.
<b>Porcentaje de aprobados por año del número de orientación vocacional</b>	Gráfico circular que muestra la cantidad y el porcentaje del promedio de los estudiantes por género en la oportunidad filtrada de la prueba específica de matemática.
<b>Cantidad de aprobados, reprobados y sin calificación</b>	Este gráfico de barras muestra el total de estudiantes que aprobaron, reprobaron y que no tuvieron calificación por género en la oportunidad filtrada de la prueba específica de matemática.

Continuación tabla XXXVII.

<b>Top de mejores notas por año del número de orientación vocacional y Género</b>	Esta gráfica de líneas muestra las mejores notas de los estudiantes por género en la prueba específica de matemática.
<b>Cantidad de estudiantes por notas</b>	Esta gráfica de área muestra la cantidad de estudiantes por notas que obtuvieron en la prueba específica de matemática, esta se encuentra dividida por el género de los estudiantes.
<b>Promedio de notas por año del número de orientación vocacional</b>	Esta gráfica de barras muestra el promedio de nota que obtuvieron los estudiantes por año del número de orientación vocacional.
<b>Estadísticos por año del número de orientación vocacional</b>	Esta tabla gráfica muestra los resultados de análisis estadísticos por el año del número de orientación vocacional.

Fuente: elaboración propia.

#### **2.4.6. Sitio web**

Se creó un sitio web para la Unidad de Análisis de Datos, el cual está diseñado para que los usuarios tengan acceso a visualizar los cuadros de mando descritos en la sección 2.4.5 y monitorear el rendimiento de los estudiantes.

Se necesita que únicamente personas que forman parte de la institución puedan acceder al sitio web, esto para que la información no sea expuesta abiertamente al público. Se definirán los usuarios y qué beneficios obtendrá cada uno dentro del sitio web.

#### **2.4.6.1. Usuario final**

Este usuario tendrá permitido visualizar los cuadros de mando del rendimiento académico, de los resultados de los estudiantes de la Facultad de Ingeniería, del curso de Matemática para Ingeniería del Programa Académico Preparatorio y de los resultados de los aspirantes a ingresar a la Facultad de Ingeniería que realizan las pruebas básicas de física y lenguaje y la prueba específica de matemática.

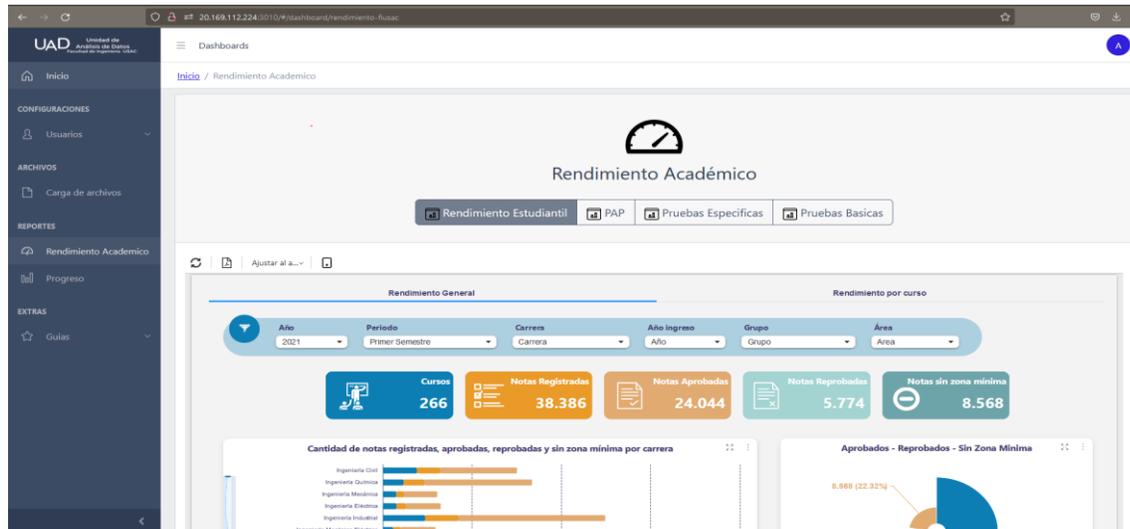
#### **2.4.6.2. Administrador**

Este usuario es quien contará con mayores privilegios que los de un usuario final. El administrador tendrá acceso a subir al sistema los archivos brindados con los datos de los resultados de los estudiantes, también tendrá permitido agregar, modificar, listar y eliminar usuarios.

#### **2.4.6.3. Incrustar cuadros de mando a sitio web**

Se tuvo la necesidad de incrustar los cuadros de mando creados con la herramienta de Pentaho Business Analysis al sitio web desarrollado con Angular. Para ello se utilizó la etiqueta *iframe*, ya que es un marco incorporado que permite insertar contenido de otras fuentes al sitio web, donde los elementos añadidos funcionan de manera autónoma.

Figura 34. **Módulo de rendimiento estudiantil**

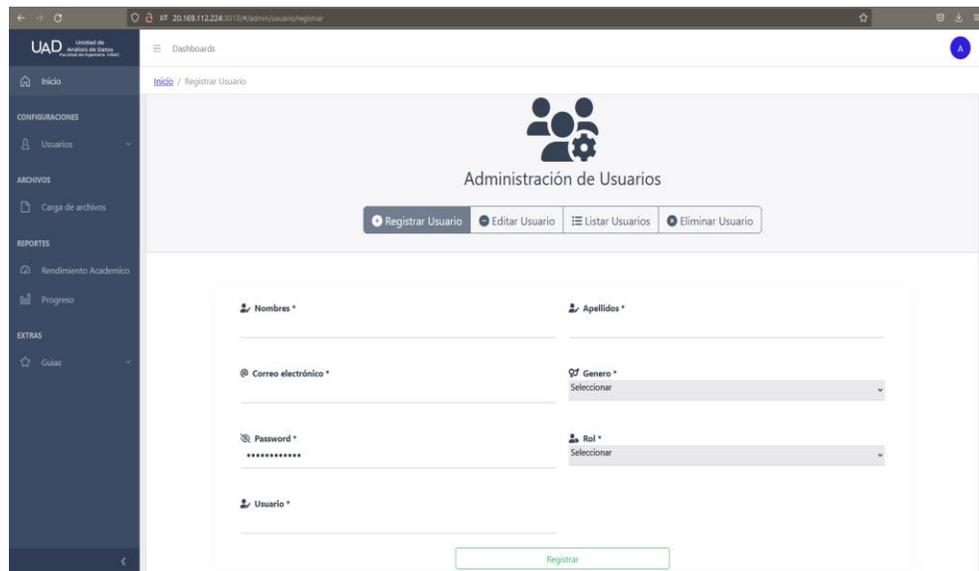


Fuente: elaboración propia.

#### 2.4.6.4. **Administración de usuarios**

Se agregó al sitio web un módulo para la administración de usuarios, a donde únicamente los usuarios con rol de administrador tendrán acceso. Este módulo permite al administrador crear, editar, listar y eliminar usuarios, de esta manera se tiene la ventaja de que solo los usuarios creados internamente tendrán acceso al sitio web, protegiendo así la información de personas que sean ajenas de la institución.

Figura 35. **Módulo de administración de usuarios**

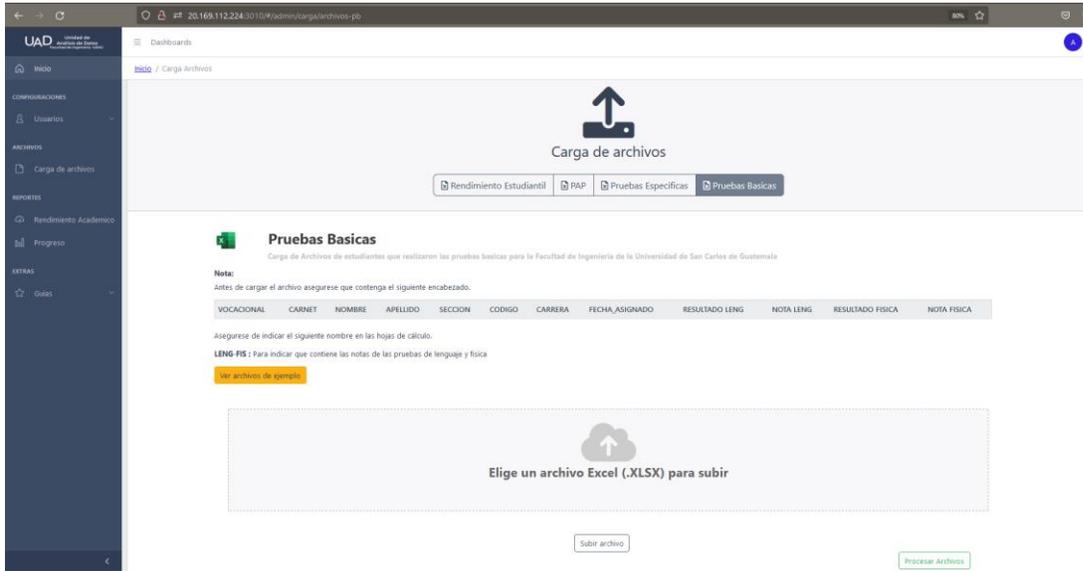


Fuente: elaboración propia.

#### 2.4.6.5. **Carga de archivos**

El sitio web cuenta con un módulo para la carga de archivos de las notas de los estudiantes, a donde únicamente los usuarios con rol de administrador tendrán acceso. Este módulo tiene apartados para subir los archivos que pertenecen a la Facultad de Ingeniería, el Programa Académico Preparatorio, las pruebas básicas y la prueba específica. El administrador tiene la opción de subir el archivo y procesar inmediatamente los datos para visualizar estos en los cuadros de mando.

Figura 36. Módulo de carga de archivos



Fuente: elaboración propia.

## 2.5. Costos del proyecto

Los costos asociados al desarrollo del proyecto se detallan en la tabla XXXVIII:

Tabla XXXVIII. Detalle de costos del proyecto

Recursos	Cantidad	Costo Unitario	Subtotal
Ingeniero de datos	1 desarrollador * 25 hrs/semana * 6 meses	Q 12,000.00 /mes	Q 72,000.00
Asesor de la Unidad de Análisis de Datos	1 asesor * 4 hrs/semana * 6 meses	Q 3,000.00 /mes	Q 18,000.00

Continuación tabla XXXVIII.

<b>Asesor de la Escuela de Ciencias y Sistemas</b>	1 asesor * 3 hrs/semana * 6 meses	Q 2,500.00 /mes	Q 15,000.00
<b>Servidor para alojamiento del sistema</b>	1 servidor * 6 meses	Q 1,000.00 /mes	Q 6,000.00
<b>Computadora</b>	1 <i>laptop</i>	Q 10,000.00	Q 10,000.00
<b>Servicio de luz</b>	1 servicio * 6 meses	Q 300.00 /mes	Q 1,800.00
<b>Servicio de Internet</b>	1 servicio * 6 meses	Q 450.00 /mes	Q 2,700.00

Total: Q 125,500.00

Fuente: elaboración propia.

## 2.6. Beneficios del proyecto

Beneficios que obtendrá la Unidad de Análisis de Datos al implementar un sistema de inteligencia de negocios para la toma de decisiones basadas en los datos:

- Analizar el comportamiento de los estudiantes en los cursos de la Facultad de Ingeniería y en el curso Matemática Preparatoria para Ingeniería del Programa Académico Preparatorio.
- Rastrear el rendimiento académico de los estudiantes de la Facultad de Ingeniería, del Programa Académico Preparatorio y de los aspirantes a entrar a la Facultad de Ingeniería que realizan las pruebas básicas y específicas.

- Predecir el éxito de los estudiantes con respecto al progreso de los cursos en el primer semestre de sus carreras dentro de la Facultad de Ingeniería.
- Identificar las tendencias de deserción en los cursos de la Facultad de Ingeniería.
- Descubrir inconvenientes en el avance de los estudiantes de la Facultad de Ingeniería en los cursos.

### **3. FASE DE ENSEÑANZA-APRENDIZAJE**

#### **3.1. Material elaborado**

Se realizaron manuales dirigidos a diversos interesados, para transmitir la información sobre la instalación del sistema, el uso correcto de este y detalles técnicos para comprender cómo fue realizada cada parte del sistema. A continuación, se describe la documentación realizada.

#### **3.2. Manual de instalación y configuración**

Se entregó un manual dirigido al responsable de instalar y configurar el sistema en el servidor de la Unidad de Análisis de Datos. Dicho manual contiene los requerimientos de hardware y de software para que el sistema funcione correctamente. Este documento describe las herramientas que conforman el sistema y para cada una se detallan los comandos necesarios para su instalación y las configuraciones necesarias.

#### **3.3. Manual técnico**

Se entregó un manual dirigido a los futuros estudiantes que continuarán agregando nuevas funcionalidades al sistema. En este se da a conocer la estructura del sistema y el uso de las herramientas. Se indica el flujo del desarrollo del proyecto, mostrando paso a paso cómo realizar los procesos de ETL, la creación de los cubos multidimensionales, el procesamiento de archivos, la creación de cuadros de mando y la integración de los cuadros de mando con el sitio web.

### **3.4. Manual de usuario**

Se entregó un manual dirigido a los usuarios del sitio web de la Unidad de Análisis de Datos, el cual contiene los pasos para el uso correcto de cada cuadro de mando. En este se encuentran descritos los reportes realizados para cada entidad y el funcionamiento de cada una, se describen los pasos para visualizar los datos en un formato plano y las opciones de descarga que tiene cada reporte.

## CONCLUSIONES

1. Se realizaron procesos de ETL acordes a las necesidades de cada entidad, y al extraer los datos de las notas de los estudiantes en los archivos recibidos se seleccionaron y limpiaron aquellos que aportan valor al análisis, para posteriormente almacenar la información en sus respectivos *Data Warehouse*.
2. Se implementó una solución de inteligencia de negocios que almacena la información y lleva un registro sobre las notas de los estudiantes a través del tiempo, en donde se realiza un análisis sobre las notas y se permite monitorear el rendimiento estudiantil para comprender el comportamiento de las notas de los estudiantes.
3. Se creó un sitio web seguro que permite a los interesados visualizar, a través de cuadros de mando, el rendimiento estudiantil de manera ordenada y hacer un seguimiento de estos datos para la toma de decisiones, para así proponer mejoras en el ámbito estudiantil al identificar comportamientos o inconvenientes.



## RECOMENDACIONES

1. Para la creación de un proceso de ETL es recomendable previamente realizar un análisis sobre las fuentes de datos para conocer de mejor manera el alcance del proyecto, así se obtendrá una idea más clara sobre los datos que aportan valor y que son relevantes para almacenar.
2. Para desarrollar una solución de inteligencia de negocios es necesario utilizar herramientas que faciliten los procesos y contar con un equipo que sea compatible con los requerimientos de software y de hardware del sistema a desarrollar. Se debe identificar las necesidades de la empresa y los datos que aportan valor para que los interesados puedan comprender el comportamiento de sus datos y medir su rendimiento.
3. Debido a la información delicada que pueden mostrar los cuadros de mando, se recomienda que el sitio web maneje un proceso de autenticación seguro, como la autenticación basada en *tokens* o la autenticación de dos pasos, los cuales garantizan que los usuarios que acceden al sitio son confiables.



## REFERENCIAS

1. Ayuware. (2022). *Cómo estandarizar datos y por qué es importante*. Recuperado de <https://www.ayuware.es/blog/como-estandarizar-datos/>
2. Bernabéu, D. (4 de enero de 2022). *Visualizer: Dashboards Mondrian OLAP en Pentaho*. LinkedIn. Recuperado de <https://es.linkedin.com/pulse/visualizer-dashboards-mondrian-olap-en-pentaho-dario-bernabeu>.
3. Bernabéu, D. y García, M. (2017). *Hefesto Data Warehousing. Guía completa de aplicación teórico-práctica. Metodología Data Warehouse*. Recuperado de [https://master.dl.sourceforge.net/project/bihefesto/Hefesto/Hefesto\\_v3.pdf?viasf=1](https://master.dl.sourceforge.net/project/bihefesto/Hefesto/Hefesto_v3.pdf?viasf=1).
4. Cupas, C. (20 de diciembre, 2022). Qué es Pentaho BI y qué módulos lo componen [Mensaje en un blog]. Recuperado de <https://openwebinars.net/blog/que-es-pentaho-bi-y-que-modulos-lo-componen/>
5. Espinosa, R. (4 de julio, 2010). Preparando el análisis dimensional. Definición de cubos utilizando Schema Workbench [Mensaje en un blog]. Recuperado de <https://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el->

analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/

6. Facultad de Ingeniería de la USAC (2022). Historia. Recuperado de <https://portal.ingenieria.usac.edu.gt/index.php/aspirante/antecedentes>
7. Nida, F. (4 de febrero, 2020). Proceso ETL y los pasos de transformación para su implementación [Mensaje en un blog]. Recuperado de <https://www.astera.com/es/type/blog/etl-process-and-steps/>
8. Proactivanet (2022). Prioridad, impacto y urgencia [Mensaje en un blog]. Recuperado de <https://www.proactivanet.com/blog/gestion-de-incidencias/prioridad-impacto-y-urgencia-1-de-2/>
9. Tableau. (s. f.). *¿Qué es la inteligencia de negocios? Guía sobre la inteligencia de negocios y por qué es importante*. Recuperado de <https://www.tableau.com/es-es/learn/articles/business-intelligence>