



**Modelo de arquitectura BI para la gestión de información de los procesos de promoción
institucional, selección y matrícula de estudiantes**

Liliana Ochoa Echeverri

Asesora

Sandra Patricia Zabala Orrego, Ingeniera Informática, Especialista en Gerencia

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería de Sistemas

Medellín

2025

Cita	Ochoa Echeverri [1]
Referencia Estilo IEEE (2020)	[1] L. Ochoa Echeverri, “Modelo de arquitectura BI para la gestión de información de los procesos de promoción institucional, selección y matrícula de estudiantes”, modalidad virtual, Ingeniería de Sistemas, Universidad de Antioquia, Medellín, 2025.



Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Cossio.

Decano/Director: Julio César Saldarriaga Molina.

Jefe departamento: Danny Alexander Múnera Ramírez.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

TABLA DE CONTENIDO

CONTENIDO

RESUMEN.....	6
ABSTRACT	7
I. INTRODUCCIÓN	8
II. PLANTEAMIENTO DEL PROBLEMA	9
III. JUSTIFICACIÓN.....	12
IV. OBJETIVOS	13
V. MARCO TEÓRICO	14
VI. METODOLOGÍA	17
VII RESULTADOS	19
VIII. DISCUSIÓN.....	25
IX. CONCLUSIONES	28
X REFERENCIAS	29

LISTA DE FIGURAS

Fig. 1. Comportamiento de matrículas 2019 – 2023. Informe de Gestión y Sostenibilidad 2023	9
Fig. 2. Comportamiento de número de inscritos por escuela y de programas con mayor número de inscritos 2024-1	21
Fig. 3. Comportamiento de número de personas que diligenciaron el formulario por escuela e inscritos por país en 2024-1	21
Fig. 4. Comportamiento del número de admitidos, por estrato, género, total de admitidos y porcentaje de no admitidos 2024-1.	22
Fig. 5. Comportamiento de la tendencia de admitidos por semestre, admitidos por nivel de estudio y admitidos por escuela 2024-1.	23
Fig. 6. Total de matriculados, porcentaje de admitidos matriculados y matriculados por estrato 2024-1	24
Fig. 7. Total de matriculados por nivel de estudios 2024-1	24

SIGLAS, ACRÓNIMOS Y ABREVIATURAS

IEEE	Institute of Electrical and Electronics Engineers
BI	Business Intelligence
AWS	Amazon, Web Services
ETL	Extracción, Transformación y Carga (Extract, Transform and Load)
EIA	Escuela de Ingeniería de Antioquia
ELT	Extracción, Carga y Transformación (Extract, Load and Transform)
IAM	Identity and Access Management
ANSI	American National Standards Institute
SQL	Structured Query Language
UdeA	Universidad de Antioquia

RESUMEN

Este proyecto tiene como objetivo optimizar los procesos de promoción, admisión y matrícula de estudiantes en una Institución de Educación Superior, mediante la centralización y organización de los datos institucionales. El enfoque principal es la creación de un Datalake en AWS S3, que centralice la información procesada y depurada, proporcionando una base de datos estructurada que permita un análisis eficiente. A través de la implementación de un flujo de trabajo ELT (Extracción, Carga y Transformación) utilizando AWS Glue, se busca automatizar la integración de datos provenientes de múltiples fuentes, garantizando su calidad, consistencia y accesibilidad. Además, la integración de AWS Quicksight para la visualización de datos facilita el análisis y la toma de decisiones informadas. Los objetivos específicos del proyecto incluyen la mejora continua de la calidad de los datos, la eliminación de duplicados y la optimización de los procesos de recolección de información, lo que resultará en la reducción de errores y reprocesos, y permitirá una toma de decisiones más estratégica y basada en datos, con el fin de impulsar la eficiencia operativa a nivel institucional.

***Palabras clave* — Centralización de datos, estudiantes, datalake, ELT, AWS glue, AWS S3, AWS Quicksight.**

ABSTRACT

This project aims to optimize the promotion, admission, and enrollment processes for students at a Higher Education Institution by centralizing and organizing institutional data. The main focus is the creation of a Data Lake in AWS S3 to centralize processed and cleaned information, providing a structured database that enables efficient analysis. Through the implementation of an ELT (Extract, Load, and Transform) workflow using AWS Glue, the goal is to automate the integration of data from multiple sources, ensuring its quality, consistency, and accessibility. Additionally, the integration of AWS Quicksight for data visualization facilitates analysis and informed decision-making. The specific objectives of the project include continuous improvement of data quality, removal of duplicates, and optimization of information collection processes, resulting in reduced errors and rework, and enabling more strategic, data-driven decision-making to enhance operational efficiency at the institutional level.

Keywords — **Data centralization, students, data lake, ELT, AWS Glue, AWS S3, AWS Quicksight.**

I. INTRODUCCIÓN

La Universidad EIA (Escuela de Ingeniería de Antioquia), ubicada en Envigado, Antioquia, Colombia, es una institución educativa fundada en 1978, reconocida por su enfoque en la excelencia académica, la investigación y el compromiso con la innovación. A lo largo de los años, la universidad ha ampliado su oferta académica, abarcando diversas áreas como ingeniería, ciencias, administración y temas relacionados con el medio ambiente y la sostenibilidad. Este crecimiento ha sido acompañado por un aumento en la población estudiantil, lo que ha generado la necesidad de adaptar sus procesos administrativos para garantizar una gestión más eficiente y alineada con las demandas actuales.

Por lo anterior, la gestión eficiente de la información en los procesos de promoción, admisión y matrícula es crucial para el buen desempeño institucional, ya que influye directamente en la toma de decisiones y en el cumplimiento de los objetivos organizacionales. Actualmente, estos procesos se ven afectados por la dispersión de la información y la falta de organización centralizada de los datos pues diferentes áreas operan de forma aislada, utilizando herramientas no integradas, como hojas de cálculo y formularios manuales, lo que da lugar a duplicaciones, errores en la información y redundancia en los procesos lo que limita la capacidad de la universidad para tomar decisiones oportunas y precisas, afectando la eficiencia operativa y la planificación a largo plazo.

Para superar estos retos, se propone la creación de un Data Lake en AWS S3, que centralice y unifique la información relevante en un único repositorio con la incorporación de herramientas avanzadas como AWS Glue para el procesamiento de datos Extracción, Carga y Transformación (ELT) y Amazon QuickSight para la visualización, este proyecto tiene como objetivo automatizar la recolección, depuración y limpieza de los datos, mejorando así su calidad y optimizando la eficiencia en la toma de decisiones. La centralización de la información permitirá a la institución contar con un sistema de inteligencia de negocios (BI) que apoye las decisiones estratégicas y operativas, contribuyendo a la modernización tecnológica y a la mejora continua en la universidad.

II. PLANTEAMIENTO DEL PROBLEMA

Dentro de los procesos esenciales para el funcionamiento del quehacer institucional de la Universidad, los procesos de promoción, admisión y matrícula de estudiantes se han caracterizado a través del tiempo como actividades críticas para el cumplimiento de los objetivos institucionales.

El comportamiento de la población estudiantil en un período de tiempo de cinco años se detalla en la Fig. 1. Y si bien, es una población que no se compara con Universidades de gran tamaño, el Plan de Desarrollo Institucional, en la visión general del año 2028, espera contar con alrededor de 3000 estudiantes matriculados. Para ello, realiza actividades de regionalización y relacionamiento activo con colegios y empresas para promover programas de pregrado, posgrado y educación continua, sin embargo, las estrategias de regionalización y posicionamiento para la atracción de talentos tienen actividad; cada área está manejando su propia información, generando que el análisis de los datos sea ineficiente.

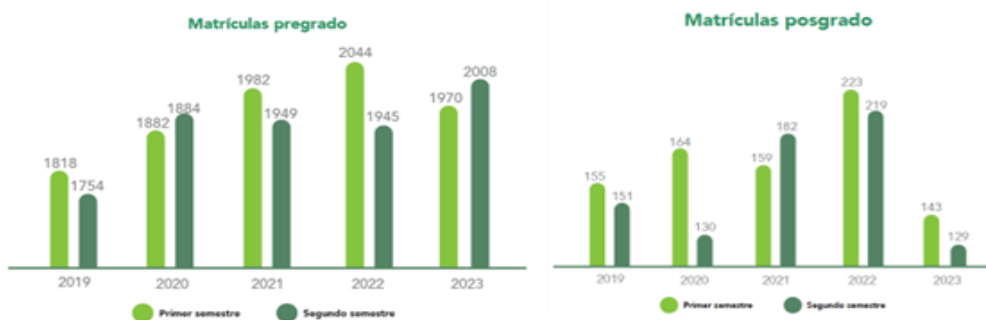


Fig. 1. Comportamiento de matrículas 2019 – 2023. Informe de Gestión y Sostenibilidad 2023

En reunión con cada proceso se logró evidenciar que existen varias fuentes de información, entre ellas, archivos de Excel con información duplicada según el área y sus necesidades y formularios físicos los cuales deben ingresar de forma manual a un sistema. Adicionalmente, no se cuenta con lineamientos institucionales para la recopilación, administración y uso de la información como por ejemplo datos personales (incluyendo información sensible).

Esta situación está causando pérdida de información, errores en los datos y reprocesos, lo que afecta negativamente la toma de decisiones y la eficiencia general de la organización.

A. Antecedentes

La integración de arquitecturas de Inteligencia de Negocios (BI) ha demostrado ser esencial en diversas organizaciones para optimizar la toma de decisiones. A continuación, se presentan varios casos de aplicación relevantes:

Universidad de Cuenca, Ecuador: La universidad enfrentaba dificultades debido a la falta de integración de sus sistemas de información, lo que afectaba la toma de decisiones. Se implementó un Data Warehouse utilizando la metodología Hefesto, que combina los enfoques de Bill Inmon y Ralph Kimball, permitiendo la definición de 7 datamarts y 27 indicadores. Herramientas como Pentaho facilitaron la visualización y el análisis de los datos, mejorando la toma de decisiones [1].

Data Warehouse para médicos tratantes de diabéticos: Se diseñó un Data Warehouse para integrar datos clínicos sobre pacientes diabéticos, utilizando un modelo entidad-relación y un ciclo de vida iterativo. Este enfoque permitió una mejora continua en el diseño y los requisitos del sistema, optimizando la toma de decisiones médicas [2].

Análisis de variables climáticas: Se desarrolló un sistema de BI para integrar datos climáticos como temperatura y humedad. Con metodologías de Bill Inmon y SCRUM, y herramientas como SQL Server, Visual Studio y Tableau, el sistema apoyó decisiones en sectores como la agricultura y la construcción [3].

Data Warehouse y minería de datos en la Universidad Nacional de la India: Se utilizó un Data Warehouse para analizar datos académicos, identificando patrones relacionados con matrícula, rendimiento y graduación. El sistema empleó un esquema de estrella y técnicas de minería de datos para facilitar el análisis académico [4].

Análisis de captación de alumnos en centros educativos: Este proyecto de BI se centró en analizar la efectividad de las campañas de marketing para la captación de estudiantes. Se diseñó un modelo lógico y físico de datos y se implementaron procesos ETL. Además, se utilizaron herramientas de gestión de proyectos como diagramas de Gantt y tableros Kanban, optimizando la planificación y ejecución del proyecto [5].

III. JUSTIFICACIÓN

La observación de las limitaciones que enfrentan muchas universidades y otras instituciones de educación superior en cuanto a la integración de sus sistemas de información, la falta de coordinación entre distintas áreas, la utilización de herramientas no integradas y la acumulación de información duplicada son factores que generan ineficiencias y errores en la gestión de los procesos académicos. Además, la dispersión de datos impide realizar un análisis adecuado que permita mejorar la calidad de la toma de decisiones. Este proyecto busca ofrecer una solución concreta a estos problemas mediante la implementación de un sistema de Business Intelligence (BI), centrado en un Datalake en AWS, que permita centralizar los datos de los procesos de promoción, admisión y matrícula en un único repositorio.

El aporte que este texto tiene a la ciencia y a la práctica está relacionado con el desarrollo e implementación de arquitecturas de BI en el ámbito educativo, en particular en el contexto de la educación superior.

A través de la propuesta de un enfoque basado en el uso de tecnologías como AWS S3, AWS Glue, Amazon Redshift y Amazon QuickSight, se busca mejorar la gestión de los datos académicos, optimizar los procesos administrativos y facilitar la toma de decisiones informadas, este enfoque, al ser aplicado en una universidad de tamaño medio, ofrece una perspectiva valiosa para otras instituciones educativas que enfrentan retos similares en la gestión de sus datos y procesos. El trabajo realizado en este proyecto no solo contribuye al ámbito académico y tecnológico, sino que también aporta al campo de la ciencia de datos y la inteligencia empresarial, al mostrar cómo las soluciones en la nube y los procesos automatizados de extracción, carga y transformación de datos (ELT) pueden ser aplicados exitosamente en entornos institucionales. De esta manera, se espera que este estudio sirva como base para futuras investigaciones y aplicaciones prácticas en otras instituciones educativas que deseen mejorar la eficiencia en la gestión de sus datos y optimizar las estrategias operativas.

IV. OBJETIVOS

A. Objetivo general

Diseñar una estructura analítica que permita centralizar y organizar la información proveniente de diferentes fuentes de información de los procesos de promoción y admisión de estudiantes de la Universidad EIA, para facilitar el análisis y la toma de decisiones.

B. Objetivos específicos

- Realizar un diagnóstico de los diferentes procedimientos y fuentes de información con la que cuenta los procesos a intervenir.
- Analizar la información almacenada y generada por el Sistema de Información Académico de la Universidad para conocer el comportamiento y el relacionamiento de estos.
- Definir la arquitectura del “datalake” en AWS S3, que centralice toda la información.
- Depurar información duplicada de los procesos, por medio de la implementación de ELT..
- Automatizar la recolección y el almacenamiento de datos: Configurar métodos automáticos de captura de datos para que no se necesitan fuentes manuales como archivos de Excel, BD externas o documentos en papel.
- Implementar un Informe en Quicksight: donde se pueda acceder y gestionar los datos desde una fuente confiable.

V. MARCO TEÓRICO

DataLake: es una arquitectura de almacenamiento de datos que permite la recopilación, almacenamiento y gestión de grandes volúmenes de datos en su formato original, ya sean estructurados, semiestructurados o no estructurados. A diferencia de un Datawarehouse, que organiza los datos de manera estructurada para el análisis inmediato, un DataLake almacena los datos de manera cruda, lo que permite una mayor flexibilidad en su procesamiento y análisis posterior [6].

En el contexto del proyecto para la Universidad EIA, el DataLake servirá como una base centralizada donde se almacenarán todos los datos relacionados con las áreas de Admisiones, Registro y Finanzas, en su formato original. Esto permitirá que los datos estén disponibles para diversas formas de análisis y procesamiento, adaptándose a las necesidades específicas de los usuarios. Además, el DataLake facilitará la integración con herramientas de visualización como QuickSight, proporcionando una infraestructura robusta para la toma de decisiones basada en datos.

Amazon Redshift: es un servicio de almacenamiento de datos en la nube, totalmente administrado y a escala de petabytes. Con Amazon Redshift, puede analizar sus datos utilizando sus herramientas de inteligencia empresarial existentes [7]. En el proyecto para la Universidad EIA, Amazon Redshift se integrará con el DataLake para facilitar el análisis de grandes volúmenes de datos de las áreas de Admisiones, Registro y Financiero.

En este enfoque, los datos se almacenarán inicialmente en el DataLake en su formato original, lo que permite flexibilidad y escalabilidad en el manejo de la información. Redshift se utilizará para realizar análisis detallados sobre los datos que requieran estructura y consultas complejas, esto permitirá aprovechar lo mejor de ambos mundos: la flexibilidad del DataLake y la capacidad analítica de Redshift.

Amazon S3: Amazon S3 es un sistema de almacenamiento para Internet. Puede usarlo para almacenar y recuperar cualquier cantidad de datos en cualquier momento y desde cualquier lugar de la web [7]. Amazon S3 será la plataforma elegida para almacenar todos los datos generados por los procesos de promoción, admisión y matrícula de la Universidad EIA, garantizando la

escalabilidad y disponibilidad de los datos para ser procesados y analizados en etapas posteriores, sin necesidad de estructurarlos inmediatamente.

AWS Lake Formation: simplifica y automatiza muchos de los pasos manuales complejos que suelen ser necesarios para crear el DataLake. Estos pasos incluyen la recopilación, limpieza, traslado y catalogación de datos, y hacer que esos datos estén disponibles de forma segura para análisis y aprendizaje automático. Lake Formation proporciona su propio modelo de permisos el: AWS Identity and Access Management (IAM) [8]. Esta herramienta permitirá automatizar tareas complejas como la recopilación, limpieza, catalogación, y la integración de los datos desde diversas fuentes, además, asegurará que los datos sean consistentes, depurados y sin duplicados antes de ser cargados en el DataLake.

AWS Glue: es un servicio de extracción, transformación y carga (ETL) completamente administrado y automatiza la ardua tarea de la preparación de datos para el análisis. AWS Glue encuentra y clasifica los datos automáticamente mediante el catálogo de datos de Glue [9]. En el contexto de este trabajo, AWS Glue, facilitará la creación de pipelines que extraigan datos desde las diferentes fuentes (bases de datos, archivos de Excel, etc.), los transformen (depuración de datos, limpieza de información duplicada) y los carguen en Amazon S3.

Amazon Quick Sight: es un servicio de análisis empresarial rápido basado en la nube que puede utilizar para crear visualizaciones, realizar análisis y obtener rápidamente información empresarial a partir de sus datos. [7].

En el contexto del DataLake para la Universidad EIA, QuickSight se utilizaría para crear tableros que consoliden los datos de Admisiones, Registro y Financiero, facilitando el análisis y la toma de decisiones. La integración directa de QuickSight con el DataLake permitirá a los administradores y directivos acceder a información actualizada en tiempo real, esencial para una gestión eficiente y estratégica. QuickSight también ofrece capacidades de machine learning que pueden ser aprovechadas para obtener insights más profundos y hacer predicciones basadas en los datos.

Athena: es un servicio de consulta interactivo que puede utilizar para analizar datos en Amazon S3 mediante ANSI SQL. Athena no tiene servidor, por lo que no hay infraestructura que administrar. Athena escala automáticamente y es fácil de usar, por lo que puede comenzar a analizar sus conjuntos de datos [9].

En este proyecto para la Universidad EIA, Athena será clave para facilitar la consulta y el análisis de los datos centralizados en Amazon S3. Al integrar Athena con otras herramientas de AWS, como QuickSight, se podrá optimizar el análisis de los datos de los procesos académicos y administrativos, apoyando una toma de decisiones informada y eficiente.

VI. METODOLOGÍA

Para resolver los problemas de dispersión y manejo ineficiente de la información en los procesos de promoción, selección y matrícula de estudiantes, se implementó una infraestructura tecnológica robusta basada en soluciones en la nube, utilizando servicios de Amazon Web Services (AWS). La infraestructura global de la solución se alojó en un centro de datos en Canadá Central, lo que permitió una alta disponibilidad y seguridad en el manejo de la información.

El primer paso en la implementación fue establecer la conexión con la base de datos de nuestro proveedor de información a través de una conexión JDBC configurada en AWS Glue. Esta conexión facilitó la extracción de las tablas necesarias para el proyecto de infraestructura. Mediante la creación de un job en AWS Glue, se realizó el proceso de extracción de datos desde la base de datos origen, y los datos extraídos fueron almacenados en un bucket de Amazon S3, en formato de archivos Parquet, que es eficiente tanto en almacenamiento como en procesamiento.

Una vez almacenados los datos en S3, se configuró un crawler en AWS Glue para realizar la validación de los datos extraídos. Este crawler se encargó de escanear los archivos Parquet y generar automáticamente el esquema correspondiente, creando la base de datos cruda dentro del Data Catalog de AWS. Este paso fue crucial para garantizar que los datos estuvieran bien organizados y fueran accesibles para su posterior análisis.

A continuación, se llevó a cabo un análisis exploratorio utilizando el módulo de Amazon Athena, lo que permitió validar los datos crudos antes de proceder con su transformación. Athena, al ser un servicio de consultas interactivas que funciona directamente sobre los datos almacenados en S3, facilitó la validación y el análisis rápido de los datos sin necesidad de moverlos a otro entorno.

Una vez validados los datos, se procedió a realizar los jobs de transformación, que fueron diseñados para limpiar y estructurar los datos de acuerdo con el modelo dimensional previamente diseñado para las bases de datos de aspiración y matrícula. Este proceso de ELT se automatizó mediante triggers configurados en AWS Glue, lo que permitió que la extracción, transformación y

carga de datos se ejecutaran de forma diaria, a las 12:00 AM, garantizando que la base de datos dimensional en Amazon Redshift estuviera siempre actualizada con los datos más recientes.

Redshift fue utilizado para alojar la base de datos dimensional, estructurada para soportar consultas rápidas y análisis detallado de los procesos de promoción, selección y matrícula. La estructura dimensional facilitó la creación de informes complejos y paneles de control.

Finalmente, la visualización de los datos y la creación de informes interactivos se realizó mediante Amazon QuickSight. Con esta herramienta, se generaron reportes personalizados y dinámicos que permitieron a las áreas encargadas de la toma de decisiones acceder a la información de manera clara y en tiempo real. Estos informes fueron compartidos de manera regular con el área responsable de la gestión de los estudiantes, facilitando la toma de decisiones informadas en cada etapa del proceso.

VII RESULTADOS

El proyecto ha logrado desarrollar una estructura analítica centralizada para los procesos de promoción y admisión de estudiantes en la Universidad EIA. Se ha obtenido un diagnóstico detallado de los procedimientos actuales y la información disponible, lo cual ha permitido realizar un análisis exhaustivo del comportamiento y relaciones de los datos en el Sistema de Información Académico. A partir de este análisis, se ha definido e implementado una arquitectura de Data Lake que organiza y centraliza la información de manera eficiente, eliminando los datos duplicados mediante un sistema ETL que ha garantizado un conjunto de datos limpio y consistente. Además, se ha automatizado la recolección y almacenamiento de datos, lo que asegura su actualización continua sin intervención manual. Finalmente, se han implementado tableros interactivos que facilitan el análisis de datos y la toma de decisiones, proporcionando a los usuarios finales herramientas visuales que les permiten extraer insights de manera ágil y precisa.

Los tableros de control en QuickSight han sido desarrollados utilizando la base de datos del sistema académico, cubriendo el período de 1999 a 2024. Estos tableros ofrecen información detallada sobre los estados de los aspirantes, los admitidos y los matriculados. A continuación, se presentan los resultados obtenidos para el semestre 2024-1.

A. Aspirantes

Según los datos mostrados en la Fig. 2, la escuela con el mayor número de aspirantes inscritos fue el Posgrado de la Escuela de Ingeniería y Ciencias Básicas, con un 51% del total, es decir, 904 estudiantes de un total de 1,766 aspirantes. En segundo lugar, se encuentra el Pregrado de la Escuela de Ciencias de la Vida y Medicina, con un 28% de los inscritos, lo que corresponde a 490 estudiantes. A continuación, en orden descendente, se encuentran:

1. Pregrado de la Escuela de Ingeniería y Ciencias Básicas, con un 12% (207 estudiantes)
2. Posgrado de la Escuela de Ciencias Económicas y Administrativas, con un 5% (80 estudiantes)
3. Pregrado de la Escuela de Ciencias Económicas y Administrativas, con un 4% (75 estudiantes).

4. Finalmente, Posgrado de la Escuela de Ciencias de la Vida y Medicina, con un 1% (10 estudiantes)

En cuanto a los programas con mayor número de aspirantes inscritos en el semestre 2024-1, los más destacados fueron, en orden descendente:

1. Especialización en Big Data e Inteligencia de Negocios
2. Medicina
3. Especialización en Inteligencia Artificial
4. Ingeniería de Sistemas y Computación
5. Ingeniería Mecatrónica

Es importante destacar que el semestre 2024-1 presentó un comportamiento atípico debido a la oferta de las "Becas 100% del Gobierno de Colombia", lo cual generó un aumento significativo en el número de aspirantes. Esta información ha sido tomada del sistema de admisiones de la Universidad EIA.

Por otro lado, en la Fig. 3, se muestra el número de personas que diligenciaron el formulario de inscripción por escuela. La escuela con mayor número de formularios diligenciados fue nuevamente el Posgrado de la Escuela de Ingeniería y Ciencias Básicas, con 1,264 inscritos, seguida por el Pregrado de la Escuela de Ciencias de la Vida y Medicina (715 inscritos), el Pregrado de la Escuela de Ingeniería y Ciencias Básicas (431 inscritos), el Posgrado de la Escuela de Ciencias Económicas y Administrativas (116 inscritos), el Pregrado de la Escuela de Ciencias Económicas y Administrativas (114 inscritos) y finalmente, el Posgrado de la Escuela de Ciencias de la Vida y Medicina (12 inscritos). Asimismo, se observa que el país con el mayor número de inscritos fue Colombia, seguido por Ecuador, Panamá y Chile.

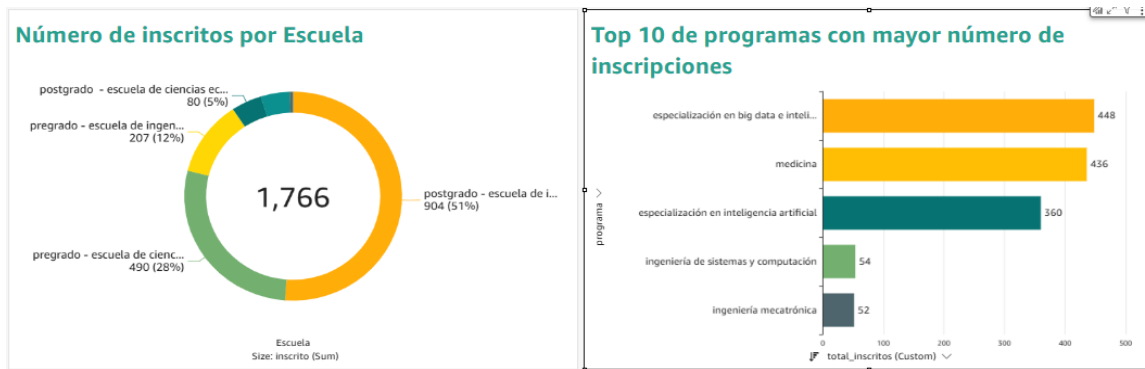


Fig. 2. Comportamiento de número de inscritos por escuela y de programas con mayor número de inscritos 2024-1



Fig. 3. Comportamiento de número de personas que diligenciaron el formulario por escuela e inscritos por país en 2024-1

B. Admitidos

En la Fig. 4, se observa que el estrato predominante entre los admitidos para el semestre 2024-1, que suman un total de 477, es el estrato 1, seguido en orden descendente por los estratos 3, 4, 6, 0, 7 y nuevamente 4. Además, se registró un 44.73% de no admitidos, lo cual se debe a que el gobierno nacional solo otorgó 2 becas para la Universidad EIA, de un total de 230 inscritos para la Especialización en Big Data e Inteligencia de Negocios. Es importante destacar que el resultado de admisión por estrato será revisado conjuntamente con el área de Admisiones, ya que el sistema

académico asigna por defecto el estrato 1, y muchos de los aspirantes no actualizan esta información durante su inscripción. En cuanto al género, se observa que la mayoría de los admitidos son de género femenino, con un 42%, seguidos por los géneros masculino (41%) y aquellos que no definieron su género, que representan un 17%.

En la Fig. 5, se puede apreciar que la tendencia de admitidos ha mostrado un aumento significativo desde el año 2023. Por nivel de estudio, la mayor parte de los admitidos son profesionales, con 370 de 477 admitidos. En cuanto al desglose por escuela, la Escuela de Ciencias de la Vida y Medicina (Pregrado) es la que concentra el mayor número de admitidos, con un 44%, seguida en orden descendente por:

1. Pregrado de la Escuela de Ingeniería y Ciencias Básicas (23%)
2. Posgrado de la Escuela de Ingeniería y Ciencias Básicas (13%)
3. Pregrado de la Escuela de Ciencias Económicas y Administrativas (11%)
4. Posgrado de la Escuela de Ciencias Económicas y Administrativas (8%)
5. Pregrado de la Escuela de Ciencias de la Vida y Medicina (2%).



Fig. 4. Comportamiento del número de admitidos, por estrato, género, total de admitidos y porcentaje de no admitidos 2024-1.



Fig. 5. Comportamiento de la tendencia de admitidos por semestre, admitidos por nivel de estudio y admitidos por escuela 2024-1.

C. Matriculados

En la Fig. 6, se muestra que de los 477 admitidos, un total de 329 estudiantes se matricularon en la Universidad EIA, lo que representa un 87.73% de matriculados respecto a los admitidos. El estrato predominante entre los matriculados, después de la normalización de los datos de estratificación, es el estrato 5, seguido por los estratos 3, 4, 6, 2 y 7.

En la Fig. 7, se observa que el nivel de estudio que predomina entre los matriculados en el semestre 2024-1 es el de Profesional, seguido por Especialización, Maestría y, finalmente, Doctorado.



Fig. 6. Total de matriculados, porcentaje de admitidos matriculados y matriculados por estrato 2024-1

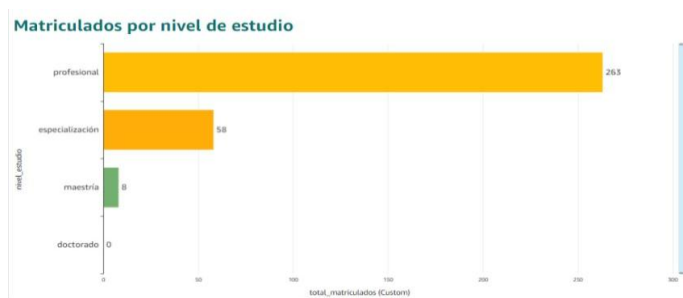


Fig. 7. Total de matriculados por nivel de estudios 2024-1

VIII. DISCUSIÓN

Los resultados obtenidos en este proyecto evidencian avances significativos en la centralización, organización y visualización de los datos de los procesos de promoción, admisión y matrícula en la Universidad EIA. A continuación, se discuten en detalle los hallazgos principales y sus implicaciones:

A. Centralización y Calidad de los Datos

La creación del Data Lake en AWS S3 y la implementación del proceso ETL (Extracción, Transformación y Carga) a través de AWS Glue han permitido centralizar la información proveniente de diversas fuentes, lo que resolvió de manera efectiva los problemas de dispersión de datos. Este enfoque ha resultado en una base de datos limpia y organizada, lo que facilita el análisis y la toma de decisiones. La eliminación de datos duplicados es especialmente relevante, ya que reduce los errores y mejora la calidad de la información, un aspecto crucial en procesos como la admisión y matrícula de estudiantes.

B. Automatización de la Recolección y Almacenamiento de Datos

Uno de los logros destacados es la automatización de la recolección y almacenamiento de datos, lo que ha permitido que los procesos sean más eficientes y menos propensos a errores humanos. Gracias a la configuración de jobs automáticos en AWS Glue, la base de datos se actualiza de manera continua sin necesidad de intervención manual. Este enfoque ha optimizado los tiempos de respuesta y ha asegurado que la información utilizada para la toma de decisiones esté siempre actualizada.

C. Visualización y Toma de Decisiones

La implementación de tableros interactivos en Amazon QuickSight ha sido fundamental para mejorar la accesibilidad de los datos. Los informes generados permiten a los responsables de

la toma de decisiones acceder a información detallada sobre los aspirantes, admitidos y matriculados, lo que facilita la identificación de tendencias, áreas de oportunidad y posibles problemas. Por ejemplo, los datos de los aspirantes permiten observar cuáles son las escuelas y programas con mayor número de inscripciones, lo que podría influir en la planificación de recursos y en la asignación de becas o plazas.

D. Comportamiento de los Aspirantes

El análisis de los datos de aspirantes reveló un comportamiento atípico en el semestre 2024-1, atribuible a la oferta de las "Becas 100% del Gobierno de Colombia". Este incremento en el número de inscripciones, especialmente en programas como la Especialización en Big Data e Inteligencia de Negocios, subraya la importancia de tener una infraestructura de datos flexible que pueda manejar fluctuaciones inesperadas en la demanda de programas académicos. Además, los datos de inscripciones por escuela y país proporcionan información valiosa para futuras campañas de promoción y estrategias de internacionalización.

E. Admitidos y Estratificación

En cuanto a los admitidos, se observa que el estrato predominante fue el estrato 1, lo que plantea una cuestión importante sobre la consistencia de los datos de estratificación en el sistema académico. La revisión conjunta con el área de Admisiones podría ayudar a resolver posibles inconsistencias en la asignación de estratos, lo que contribuiría a una gestión más precisa y equitativa de las becas y los recursos. Además, el hecho de que la mayoría de los admitidos sean mujeres (42%) es un dato relevante para futuras iniciativas de inclusión y equidad de género dentro de la universidad.

F. Matriculados

De los 477 admitidos, un 87.73% logró completar el proceso de matrícula, lo que indica un alto índice de conversión de admitidos a matriculados. Este alto porcentaje refleja la efectividad de los procesos de orientación y apoyo a los estudiantes, aunque también sugiere que la universidad

debe seguir evaluando las razones por las que algunos admitidos no concretan su matrícula, lo que podría estar relacionado con factores como la oferta económica, el apoyo logístico o la percepción de la calidad educativa.

G. Tendencias y Demanda de Programas

En el análisis de los matriculados por programa y nivel de estudio, se observa una tendencia hacia el incremento de la matrícula en programas profesionales, con un 87.73% de matriculados en este nivel. Este patrón, junto con el predominio de las especializaciones y maestrías, podría sugerir una creciente demanda de formación avanzada en áreas específicas, lo cual podría guiar las decisiones de la universidad en cuanto a la expansión de programas de posgrado. Además, el comportamiento por estrato muestra que la universidad podría explorar nuevas estrategias para atraer estudiantes de diferentes estratos socioeconómicos, especialmente en estratos más altos.

H. Impacto de las Becas y Políticas Gubernamentales

La baja tasa de admisión en la Especialización en Big Data e Inteligencia de Negocios debido a la limitación de becas disponibles ilustra cómo las políticas gubernamentales y la disponibilidad de financiamiento pueden impactar directamente en las decisiones de los estudiantes. Este tipo de información es esencial para que la universidad pueda negociar con entidades gubernamentales o privados para ampliar la oferta de becas o desarrollar alternativas de financiamiento para los aspirantes.

IX. CONCLUSIONES

Usando AWS Glue para ETL, Redshift para almacenamiento dimensional y Athena para análisis, se garantiza la validación y procesamiento correcto de los datos, además, la automatización con triggers y la visualización en QuickSight facilitan informes precisos y en tiempo real, mejorando la toma de decisiones y la eficiencia operativa y ofreciendo flexibilidad y escalabilidad para futuras necesidades de análisis.

Implementar una arquitectura de BI en las empresas es crucial para centralizar los procesos y optimizar la gestión de datos, permitiendo una toma de decisiones más rápida y precisa.

Al adoptar un enfoque de Data Lake en una plataforma de servicios en la nube como AWS, y aplicar un proceso de ELT, las empresas tienen la capacidad de consolidar grandes volúmenes de datos en un repositorio centralizado.

El uso del enfoque ELT mejora significativamente la eficiencia en el manejo de grandes volúmenes de datos al permitir su centralización sin necesidad de preprocesarlos antes de la carga. Este modelo facilita el almacenamiento de los datos en su formato más crudo, lo que ofrece mayor flexibilidad para su análisis y procesamiento posterior según las necesidades del negocio.

La depuración de datos duplicados y la implementación de un sistema ELT han asegurado que los datos almacenados sean consistentes y limpios, lo cual es esencial para realizar análisis confiables.

X REFERENCIAS

- [1] J. Muñoz y C. Galarza, "Data Warehouse para la Universidad de Cuenca: Metodología Hefesto y Visualización de Datos con Pentaho," 2017.
- [2] J. Llorente, et al., "Diseño de un Data Warehouse para médicos tratantes de diabéticos," 2012.
- [3] S. Quishpe, et al., "Sistema de análisis de variables climáticas con herramientas de BI," 2020.
- [4] L. Edastama, et al., "Data Warehouse y minería de datos en la Universidad Nacional de la India," 2021.
- [5] D. Roperó Alcazar y M. Martínez Fonte, "Análisis de captación de alumnos en centros educativos utilizando BI," 2021.
- [6] J. C. Agudelo Patiño, "Data Lakes: Aplicaciones, Herramientas y Arquitecturas," 2020. [En línea]. Disponible en: <https://repositorio.utp.edu.co/server/api/core/bitstreams/5f56e572-d416-487e-a6d5-ec3a8e45da46/content>. [Accedido: 20-ene-2025].
- [7] Amazon Web Services, "Glosario de AWS," AWS Documentation, 2025. [Enlace: <https://docs.aws.amazon.com/glossary/latest/reference/glos-chap.html>]. [Accedido: 20-ene-2025].
- [8] T. Zúñiga y K. Francini, "Diseñar un data lake en aws lake formation para mejorar el almacenamiento y el acceso de los datos," [Enlace: https://repositorio.ucenfotec.ac.cr/bitstream/handle/123456789/350/Karina%20Tencio%20Z%c3%ba%c3%bliga-MBD_Octubre%202021.pdf?sequence=1&isAllowed=y], 2021. [Accedido: 20-ene-2025], p. 36.
- [9] T. Zúñiga y K. Francini, "Diseñar un data lake en aws lake formation para mejorar el almacenamiento y el acceso de los datos," [Enlace: https://repositorio.ucenfotec.ac.cr/bitstream/handle/123456789/350/Karina%20Tencio%20Z%c3%ba%c3%bliga-MBD_Octubre%202021.pdf?sequence=1&isAllowed=y], 2021. [Accedido: 20-ene-2025], p. 37.