



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL – SNIES 2828

TALLER ETL - “EDUSMART”

Estudiante

LAURA VALENTINA BUSTOS LUPACO

PROGRAMA INGENIERÍA INDUSTRIAL

CIENCIA DE DATOS

Profesor

JESUS ARIEL GONZALEZ BONILLA

3 Octubre 2025



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito



Documentación del Flujo ETL EduSmart

Introducción

En la actualidad, las instituciones educativas generan grandes volúmenes de información relacionados con sus estudiantes, cursos y matrículas. Sin embargo, estos datos suelen encontrarse en diferentes formatos y sin un proceso de organización adecuado, lo que dificulta su análisis.

El objetivo de este proyecto es diseñar y ejecutar un flujo ETL (Extracción, Transformación y Carga) que permita centralizar y depurar los datos de la empresa ficticia EduSmart, garantizando información confiable para la toma de decisiones académicas.

Justificación

La implementación de un flujo ETL es fundamental en el área de ciencia de datos, ya que asegura la calidad, consistencia e integridad de la información antes de que esta sea utilizada en procesos de análisis o inteligencia de negocios.

En el caso de EduSmart, un correcto procesamiento de datos permitirá:

- Identificar de manera precisa el estado de las matrículas.
- Garantizar que la información de los estudiantes esté libre de duplicados y errores.
- Calcular métricas clave como el costo total de los cursos en función de su duración.
- Contar con una base de datos unificada que soporte futuros análisis académicos y financieros.

Escenario

EduSmart es una plataforma de educación en línea que ofrece cursos de programación, ciencia de datos y tecnología. La información disponible se encuentra en un único archivo JSON que contiene tres secciones principales:

- Estudiantes: registros con nombre y correo de los inscritos.
- Cursos: información de cada curso, incluyendo costo y duración en semanas.
- Matrículas: datos de inscripción de estudiantes en cursos específicos con fecha asociada.

El objetivo del flujo ETL es transformar estos datos en un formato limpio y estructurado, almacenado en una tabla final llamada `matriculas_limpias` en una base de datos SQLite.



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL – SNIES 2828

Transformación de Datos

Durante la fase de transformación se aplicaron las siguientes reglas de calidad:

1. Estudiantes

- Eliminación de valores nulos.
- Eliminación de registros duplicados.
- Normalización de nombres en mayúsculas.

2. Matrículas

- Conversión de las fechas a un formato estándar ISO (YYYY-MM-DD).
- Validación de claves foráneas para garantizar que cada matrícula corresponda a un estudiante y un curso existente.

3. Cursos

- Validación de que los costos y la duración de los cursos fueran mayores a cero.

4. Integración final

- Unión de estudiantes, cursos y matrículas en una única tabla.
- Creación de la columna calculada $\text{costo_total} = \text{costo} * \text{duracion_semanas}$, como indicador del valor académico y financiero de cada matrícula.

📍 Sede Quirinal: Calle 21 No. 6 - 01

📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220

📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699

✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

INSTITUCIÓN DE EDUCACIÓN SUPERIOR SUJETA A INSPECCIÓN
Y VIGILANCIA POR EL MINISTERIO DE EDUCACIÓN NACIONAL - SNIES 2828

Conclusión

El desarrollo del flujo ETL para EduSmart demostró que incluso partiendo de una única fuente de datos (JSON), es indispensable aplicar un proceso estructurado de limpieza y normalización para garantizar la calidad de la información.

Gracias a las transformaciones aplicadas, los datos ahora se encuentran listos para análisis académicos y financieros, permitiendo identificar patrones de matrícula, ingresos proyectados y calidad de la base estudiantil.

📍 Sede Quirinal: Calle 21 No. 6 - 01
📍 Sede Prado Alto: Calle 8 No. 32 - 49 PBX: (608) 8754220
📍 Sede Pitalito: Carrera 2 No. 1 - 27 - PBX: (608) 8360699
✉ Email: contacto@corhuila.edu.co - www.corhuila.edu.co

Personería Jurídica Res. Ministerio de Educación No. 21000 de Diciembre 22 de 1989
NIT. 800.107584-2



CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA

Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito