

**DISEÑO E IMPLEMENTACIÓN DE UNA APLICACIÓN WEB PARA LA
GESTIÓN Y TRAZABILIDAD DE ACTAS DE SALIDA DE MATERIALES EN EL
ÁREA DE INSPECCIONES DE ENEL COLOMBIA**

ANGIE DANIELA GONZALEZ ESPAÑA

**UNIVERSIDAD SANTO TOMÁS
FACULTAD DE INGENIERÍA
INGENIERÍA DE TELECOMUNICACIONES
BOGOTÁ D.C**

2026

**DISEÑO E IMPLEMENTACIÓN DE UNA APLICACIÓN WEB PARA LA
GESTIÓN Y TRAZABILIDAD DE ACTAS DE SALIDA DE MATERIALES EN EL
ÁREA DE INSPECCIONES DE ENEL COLOMBIA**

Presentado por:

ANGIE DANIELA GONZALEZ ESPAÑA

CÓDIGO: 2326166

TRABAJO OPCIÓN DE GRADO PASANTÍAS EN LA EMPRESA ENEL.

DIRECTOR:

VÍCTOR MANUEL CASTRO

INGENIERO ELECTRÓNICA

UNIVERSIDAD SANTO TOMÁS

FACULTAD DE INGENIERÍA

INGENIERÍA DE TELECOMUNICACIONES

BOGOTÁ D.C

2026

Nota de aceptación:

Firma Tutor

Firma Jurado

Firma Jurado

Bogotá 24 de junio del 2026.

TABLA DE CONTENIDO

INTRODUCCIÓN	9
1. PLANTEAMIENTO DEL PROBLEMA	11
2. OBJETIVOS	13
2.1. Objetivo general	13
2.2. Objetivos específicos.....	13
3. METODOLOGÍA	14
3.1. Etapa 1: Análisis del proceso actual.....	14
3.2. Etapa 2: Diseño de la solución	14
3.3. Etapa 3: Desarrollo e implementación.....	15
3.4. Etapa 4: Validación de la solución.....	15
3.5. Requerimientos del sistema	15
3.6. Criterios de aceptación	17
4. MARCO TEÓRICO	18
4.1. Gestión de la información	18
4.2. Procesos ETL (Extracción, Transformación y Carga)	18
4.3. Calidad de los datos	18
4.4. Bases de datos y centralización de la información	19
4.5. Automatización de procesos	19
4.6. Herramientas tecnológicas utilizadas	19
5. DESARROLLO	21
5.1. Etapa 1: Análisis del proceso actual.....	21
5.2. Etapa 2: Diseño de la solución	24
5.3. Etapa 3: Desarrollo e implementación	27
5.3.1. Etapa 3.1. Arquitectura y diseño del sistema	27
5.3.2. Etapa 3.2. Desarrollo del módulo de procesamiento de datos	31
5.3.3. Etapa 3.3. Integración con la base de datos	34
5.3.4. Etapa 3.4. Desarrollo de la interfaz de usuario	35
5.3.5. Etapa 3.5. Implementación del módulo de carga	37
5.4. Etapa 4. Validación de la solución	38

5.4.1. Etapa 4.1. Metodología de validación	39
5.4.2. Etapa 4.2. Análisis de indicadores	40
5.4.3. Etapa 4.3. Resultados de la validación.....	41
CONCLUSIONES	44
REFERENCIAS	46

LISTA DE FIGURAS

Figura 1: Plataforma interna de logística donde se publican los archivos de despacho de materiales.	21
Figura 2: Ejemplo de estructura de archivo AS con fórmula BUSCARV en columna irrelevante.....	22
Figura 3: Ejemplo de estructura de archivo AS con inconsistencias en los datos y columnas no relevantes.....	22
Figura 4: Flujo general del proceso de transformación, almacenamiento y consulta de la información propuesto en la solución	25
Figura 5: Diagrama de flujo del proceso de carga, procesamiento y almacenamiento de datos	28
Figura 6: Diagrama de clases del sistema de gestión de materiales.....	29
Figura 7: Estructura y contenido de la tabla gm_recovery.materiales en PostgreSQL, donde se almacenan los registros de materiales seriados procesados a partir de los archivos AS.	31
Figura 8: Detección automática de encabezados en archivos Excel	31
Figura 9: Lectura del archivo con encabezado identificado.....	32
Figura 10: Estandarización de columnas	32
Figura 11: Eliminación de registros sin información relevante	33
Figura 12: Generación de la referencia	33
Figura 13: Datos procesados y normalizados tras la etapa de limpieza.....	33
Figura 14: Datos procesados y normalizados tras la etapa de limpieza (sellos)	34
Figura 15: Inserción de datos en PostgreSQL.....	34
Figura 16: Registro y control de archivos procesados	35
Figura 17: Tabla archivos procesados en PostgreSQL.....	35
Figura 18: Interfaz de búsqueda de materiales.....	36
Figura 19: Resultado de búsqueda de materiales en la interfaz	36
Figura 20: Módulo de carga de archivos.....	37
Figura 21: Proceso de carga masiva.....	38
Figura 22: Ejecución del proceso de carga masiva de archivos.....	38
Figura 23: Comparación de indicadores antes y después de la implementación	43

LISTA DE TABLAS

Tabla 1: Requerimientos funcionales y no funcionales del sistema.....	17
Tabla 2: Criterios de aceptación del sistema.....	17
Tabla 3: Comparación cuantitativa de indicadores antes y después de la implementación	42

LISTA DE MATRICES

Matriz 1: Selección de framework para la interfaz de usuario.....	26
Matriz 2: Selección del motor de base de datos	26

INTRODUCCIÓN

En el contexto actual de las organizaciones, el manejo adecuado de la información se ha convertido en un elemento clave para la toma de decisiones y la optimización de los procesos operativos. En empresas del sector energético como Enel Colombia, donde se gestionan grandes volúmenes de datos asociados a activos técnicos, resulta fundamental contar con herramientas que permitan organizar, analizar y consultar la información de manera eficiente.

Enel es una empresa multinacional del sector energético con presencia en más de 30 países, reconocida como uno de los principales operadores integrados en los mercados de electricidad y gas a nivel mundial. Su presencia en Colombia tiene origen en los procesos de transformación del sector energético nacional, destacándose la evolución de la Empresa de Energía de Bogotá (EEB) y la posterior creación de Enel-Codensa en 1997, tras la participación accionaria del grupo Enel en el mercado colombiano. Desde entonces, la compañía ha consolidado su operación en el país, desempeñando un papel fundamental en la distribución y comercialización de energía eléctrica en Bogotá y Cundinamarca.

Dentro de su estructura organizacional, Enel se divide en diferentes líneas de negocio como Grids, Green Power y Staff. En el marco de la presente práctica profesional, el proyecto se desarrolló en la línea de Grids, específicamente en el área de Inspecciones Colombia. Esta área es responsable de la verificación, control y trazabilidad de activos técnicos como medidores, transformadores de corriente (CT) y sellos de seguridad, los cuales son entregados por el área de Logística a diferentes empresas contratistas. Estas entregas se realizan a través de Puntos de Despacho de Inventario (PDI), que corresponden a los puntos o contratistas receptores de los materiales, entre los cuales se encuentran OCA, INMEL y APPLUS. Cada entrega realizada a un PDI queda registrada en un documento denominado Acta de Salida (AS), un archivo en formato Excel que contiene el detalle de los materiales seriados despachados. Son precisamente estos archivos AS los que el área de Inspecciones debe consultar y analizar para verificar la trazabilidad de los activos entregados.

Dentro de este entorno, el área de Inspecciones desempeña un papel importante en la verificación y control de los materiales entregados a los diferentes PDIs. Sin embargo, la información contenida en los AS se encuentra distribuida en múltiples archivos Excel con estructuras poco estandarizadas, lo que dificulta su análisis y genera una alta carga operativa en los procesos de verificación. Esta situación se ve agravada por el volumen de Actas de Salida generadas periódicamente, lo que hace inviable sostener el proceso de verificación de manera completamente manual sin incurrir en errores, retrasos o riesgos de pérdida de información.

A partir de esta problemática surge la necesidad de implementar una solución que permita optimizar el manejo de la información, reduciendo los tiempos de procesamiento y mejorando la trazabilidad de los materiales. En este sentido, el presente proyecto tiene como objetivo desarrollar una aplicación web que automatice el procesamiento de los archivos AS, permita su consolidación en una base de datos centralizada y facilite su consulta a través de una interfaz amigable.

Para el desarrollo de la solución se utilizaron herramientas como Python, PostgreSQL y Streamlit, las cuales permitieron implementar un proceso automatizado de extracción, transformación y carga de datos (ETL), así como la construcción de una aplicación que responde a las necesidades del área de Inspecciones.

Vale la pena precisar qué existía en el área antes del desarrollo de este proyecto y cuáles fueron los aportes concretos realizados durante la práctica. Al momento de iniciar, el área de Inspecciones contaba con una base de datos PostgreSQL en uso, con su propia estructura de tablas operativas, y con un proceso de verificación de materiales que se realizaba manualmente en Microsoft Access, tarea que históricamente había recaído en los practicantes del área sin que se hubiera logrado consolidar una solución permanente. A partir de ese contexto, los aportes desarrollados en este proyecto fueron: el diseño e implementación del mecanismo de detección dinámica de encabezados para archivos Excel con estructuras no estandarizadas; el diseño del proceso ETL completo, incluyendo extracción desde múltiples formatos, transformación con normalización de columnas, limpieza de registros, generación del campo de referencia y extracción del número de AS y contrato desde el nombre del archivo; el diseño del esquema de dos tablas nuevas integradas en la base de datos existente del área; la construcción de la interfaz web con Streamlit, con módulo de búsqueda y módulo de carga manual y masiva; y la definición del mecanismo de control de duplicados mediante la tabla de archivos procesados y la instrucción `ON CONFLICT DO NOTHING`.

El presente documento se organiza en cuatro secciones principales: el planteamiento del problema, que describe el contexto y las limitaciones del proceso actual; los objetivos, que definen el alcance del proyecto; el desarrollo, que detalla cada etapa de construcción e implementación de la solución; y la validación, que presenta los resultados obtenidos y su impacto en el proceso del área de Inspecciones.

1. PLANTEAMIENTO DEL PROBLEMA

En el contexto operativo de Enel Colombia, el área de Inspecciones desempeña un papel fundamental en la verificación y control de los activos técnicos utilizados en la infraestructura eléctrica, tales como medidores, transformadores de corriente (CT) y sellos de seguridad.

Dentro de este proceso, existe una interacción constante con el área de Logística, la cual se encarga de gestionar y registrar la entrega de materiales a los diferentes Puntos de Despacho de Inventario (PDI), que corresponden a los contratistas receptores de los materiales, entre los que se encuentran OCA, INMEL y APPLUS. Cada entrega realizada a un PDI queda registrada en un Acta de Salida (AS), que es un archivo en formato Excel generado por el área de Logística y publicado en una plataforma interna. Este archivo contiene el detalle de los materiales seriados entregados, es decir, cada uno de los activos identificados mediante un serial único, como medidores o transformadores de corriente. Son precisamente estos archivos AS los que el área de Inspecciones debe consultar para verificar si un material fue efectivamente entregado a un PDI determinado.

A pesar de la existencia de esta plataforma, el modelo actual de gestión de la información presenta múltiples limitaciones que afectan directamente la eficiencia operativa del área de Inspecciones.

En primer lugar, se evidencia una alta fragmentación de la información, ya que los datos se encuentran distribuidos en múltiples archivos AS, cada uno compuesto por diversas hojas con estructuras heterogéneas. Esta falta de estandarización impide consolidar fácilmente la información y dificulta la trazabilidad de los materiales a lo largo del proceso.

Adicionalmente, se presentan inconsistencias en la calidad del dato, derivadas principalmente del registro manual realizado por el área de Logística. Entre los problemas más frecuentes se encuentran:

- Seriales incompletos o con formatos inconsistentes
- Columnas irrelevantes para el área de Inspecciones
- Información clave como el número de AS y el contrato asociado, contenida únicamente en el nombre del archivo, sin estar registrada dentro de la tabla de datos

En cuanto al nombre de los archivos, estos siguen en su mayoría un formato estándar definido por el área de Logística, como, por ejemplo: 01_DESPACHOS_01_ARCHIVOS_DH2004319_APPLUS_NORCONTROL_INSPECCIONES_AS12843_ARC HIVOS_OPERACIÓN_AS12843.LISTADO_SERIALES.113621. Sin embargo, en casos en los que los archivos son enviados directamente por correo electrónico u otras vías alternas, dicho formato puede variar, lo que representa una dificultad adicional para la extracción

automatizada de esta información. Estas inconsistencias obligan a realizar procesos adicionales de limpieza y transformación antes de poder utilizar la información de manera efectiva.

Como consecuencia de lo anterior, cuando un contrato reporta la no recepción de materiales, el equipo de Inspecciones debe llevar a cabo un proceso manual de verificación, el cual consiste en la descarga de múltiples archivos AS desde la plataforma interna o desde correos electrónicos, la revisión hoja por hoja y la consolidación manual de la información relevante.

Este procedimiento no solo resulta altamente tedioso, sino que también representa una alta carga operativa, ya que la construcción de una base de datos para un solo contrato puede tomar entre dos y tres días de trabajo continuo. Durante este proceso, es necesario realizar tareas repetitivas como la selección de columnas, la generación de referencias a partir de los seriales y la identificación del contrato y del AS correspondiente a partir del nombre del archivo, lo que incrementa significativamente la probabilidad de errores humanos.

Finalmente, esta situación genera una limitación en la capacidad de respuesta del área, así como una vulnerabilidad en el control de los activos, ya que la dificultad para verificar de manera ágil si un material fue efectivamente entregado por Logística puede derivar en pérdidas de inventario, inconsistencias operativas e incluso riesgos asociados a posibles fraudes.

En este sentido, surge la necesidad de diseñar e implementar una solución tecnológica que permita automatizar el procesamiento de los archivos AS, garantizar la calidad de la información y centralizar los datos en una estructura que facilite su consulta, análisis y trazabilidad.

A partir de lo anterior, se plantea la siguiente pregunta de investigación: ¿Cómo automatizar el procesamiento, consolidación y consulta de la información contenida en los archivos AS del área de Inspecciones de Enel Colombia, con el fin de reducir los tiempos de verificación y mejorar la trazabilidad de los materiales entregados a los diferentes PDIs?

2. OBJETIVOS

2.1. Objetivo general

Desarrollar una aplicación web basada en Python y PostgreSQL para la automatización del procesamiento, almacenamiento y consulta de los archivos de Actas de Salida (AS) del área de Inspecciones de Enel Colombia, con el fin de mejorar la trazabilidad de los activos técnicos entregados a los diferentes Puntos de Despacho de Inventario (PDI), reducir los tiempos de verificación y fortalecer los mecanismos de control operativo.

2.2. Objetivos específicos

- Diseñar un mecanismo de lectura y detección automática de encabezados en archivos Excel y CSV con estructuras no estandarizadas, que identifique y extraiga la información relevante sin intervención manual.
- Implementar un proceso de limpieza y transformación de datos que normalice la información proveniente de múltiples archivos AS, mediante la corrección de inconsistencias, la eliminación de registros incompletos, la generación de referencias a partir de los seriales y la extracción del número de AS y el contrato asociado a partir del nombre del archivo.
- Diseñar e implementar el esquema de almacenamiento en la base de datos PostgreSQL del área de Inspecciones, desarrollando un mecanismo de carga masiva que evite duplicidades, garantice la integridad de los registros consolidados y facilite su consulta centralizada.
- Desarrollar una interfaz de consulta interactiva que permita realizar búsquedas sobre los datos almacenados mediante criterios como serial, referencia o código, facilitando la verificación de la entrega de materiales y la toma de decisiones operativas dentro del área de Inspecciones.
- Validar el funcionamiento del sistema mediante pruebas con datos reales del área de Inspecciones, comparando indicadores clave de desempeño antes y después de la implementación, con el fin de evidenciar el impacto operativo de la solución desarrollada.

3. METODOLOGÍA

El presente proyecto corresponde a un desarrollo tecnológico aplicado, estructurado a partir de un ciclo de ingeniería que comprende el diagnóstico del proceso actual, el diseño de la solución, la implementación por módulos, las pruebas con datos reales y la validación mediante indicadores de desempeño. Este enfoque responde a la naturaleza del problema abordado: una necesidad operativa concreta dentro del área de Inspecciones de Enel Colombia, resuelta mediante el diseño e implementación de una herramienta de software funcional.

Dentro de este marco, el proyecto adopta un enfoque cuantitativo en su etapa de validación, dado que se miden variables como el tiempo de procesamiento y la cantidad de registros gestionados, y un alcance descriptivo y propositivo, ya que parte del análisis de la situación actual para plantear una solución que responda a las necesidades identificadas. El conocimiento del proceso actual fue obtenido a partir de la experiencia directa adquirida durante la práctica profesional, en la cual la autora se desempeñó como apoyo al encargado de materiales y sistemas del área de Inspecciones, participando activamente en las actividades de verificación y consolidación de los archivos AS. Adicionalmente, el encargado de materiales y sistemas confirmó que esta misma dificultad había sido enfrentada de manera recurrente tanto por él como por los practicantes anteriores del área. Esta inmersión directa en el proceso eliminó la necesidad de entrevistas formales adicionales, dado el nivel de involucramiento de la autora en las tareas afectadas.

Para la ejecución del proyecto se definieron las siguientes etapas:

3.1. Etapa 1: Análisis del proceso actual

En esta etapa se realizó un diagnóstico del flujo de trabajo del área de Inspecciones, identificando las actividades involucradas en la verificación de la entrega de materiales y las principales dificultades asociadas al manejo de la información.

3.2. Etapa 2: Diseño de la solución

Con base en los problemas identificados, se definió la estructura general de la solución, estableciendo las herramientas y componentes necesarios para el procesamiento, organización y almacenamiento de los datos. La selección de herramientas se realizó con base en criterios de compatibilidad con la infraestructura existente del área, facilidad de implementación y capacidad de procesamiento requerida. Se seleccionó Python como lenguaje principal por su versatilidad en el manejo de datos, PostgreSQL por ser la plataforma de base de datos ya utilizada por el área de Inspecciones, y Streamlit por su capacidad para desarrollar interfaces web de manera rápida y accesible. En esta etapa también se definió el

esquema de las dos tablas nuevas que serían creadas dentro de la base de datos existente: una para el almacenamiento de los registros de materiales y otra para el control de los archivos AS procesados.

3.3. Etapa 3: Desarrollo e implementación

Se desarrolló la solución tecnológica, incluyendo los procesos de lectura automática de archivos, detección dinámica de encabezados, limpieza y transformación de datos, integración con la base de datos mediante carga masiva, y construcción de la interfaz de usuario para la consulta de la información. Cada uno de estos componentes fue desarrollado de manera modular, lo que permitió probar y ajustar cada parte de forma independiente antes de integrarlas en el sistema final.

3.4. Etapa 4: Validación de la solución

Se realizaron pruebas con datos reales provenientes del área de Logística, procesando un total de 21 archivos AS correspondientes a diferentes PDIs, lo que resultó en la consolidación de más de 15.000 registros de materiales seriados. Estos mismos datos habían sido procesados previamente de manera manual utilizando Microsoft Access, mediante un proceso que implicaba la importación individual de cada hoja de cálculo, la selección manual de columnas relevantes, la adición de las columnas de contrato y AS en cada hoja, y la consolidación mediante consultas de anexo. Este proceso manual permitió establecer una línea base de comparación para evaluar el impacto real de la solución implementada en términos de tiempo de procesamiento, carga operativa y calidad de los datos obtenidos.

3.5. Requerimientos del sistema

Con el fin de orientar el desarrollo de la solución, se definieron los requerimientos funcionales y no funcionales del sistema, los cuales establecen las capacidades esperadas de la aplicación y las condiciones de calidad bajo las cuales debe operar.

ID	Tipo	Requerimiento	Descripción
RF – 01	Funcional	Procesamiento automático de archivos AS	El sistema debe leer y procesar archivos en formato Excel (.xlsx) y CSV sin intervención manual en cada paso.
RF – 02	Funcional	Detección dinámica de encabezados	El sistema debe identificar automáticamente la fila de encabezado en archivos con estructuras no estandarizadas, sin requerir configuración manual por archivo.
RF – 03	Funcional	Limpieza y normalización de datos	El sistema debe eliminar registros sin serial, normalizar nombres de

			columnas y descartar columnas irrelevantes o duplicadas.
RF – 04	Funcional	Generación del campo de referencia	El sistema debe generar automáticamente el campo referencia a partir de los últimos ocho caracteres del serial de cada material.
RF – 05	Funcional	Extracción de metadatos desde el nombre del archivo	El sistema debe extraer el número de AS y el contrato asociado directamente desde el nombre del archivo AS.
RF – 06	Funcional	Carga masiva a PostgreSQL	El sistema debe insertar los registros procesados en la base de datos mediante carga masiva optimizada.
RF – 07	Funcional	Control de duplicados	El sistema debe detectar archivos ya procesados y omitirlos del proceso de carga, evitando registros duplicados en la base de datos.
RF – 08	Funcional	Búsqueda de materiales	El sistema debe permitir consultar materiales por serial completo, referencia o código a través de la interfaz de usuario.
RF – 09	Funcional	Carga manual y masiva de archivos	El sistema debe ofrecer dos modalidades de carga: archivo por archivo mediante interfaz gráfica, y carga masiva desde una carpeta local.
RNF – 01	No funcional	Tiempo de respuesta en búsquedas	Las consultas sobre la base de datos deben retornar resultados en menos de 5 segundos.
RNF – 02	No funcional	Capacidad de procesamiento	El sistema debe ser capaz de procesar más de 15.000 registros en una sola ejecución sin degradación del rendimiento.
RNF – 03	No funcional	Compatibilidad de formatos	El sistema debe soportar archivos en formato .xlsx y .csv, independientemente de su estructura interna.
RNF – 04	No funcional	Usabilidad	La interfaz debe ser operable por usuarios sin conocimientos técnicos avanzados en programación o bases de datos.

RNF – 05	No funcional	Integridad de datos ante fallos	En caso de error durante la inserción, el sistema debe ejecutar un rollback que garantice que no queden datos parcialmente cargados en la base de datos.
-----------------	--------------	---------------------------------	--

Tabla 1: Requerimientos funcionales y no funcionales del sistema

3.6. Criterios de aceptación

Para determinar el correcto funcionamiento del sistema antes de su validación formal, se definieron los siguientes criterios de aceptación, los cuales establecen las condiciones mínimas que la solución debía cumplir para considerarse exitosa:

ID	Criterio	Condición de aceptación
CA – 01	Tiempo de procesamiento	El sistema debe procesar los 21 archivos AS con más de 15.000 registros en un tiempo significativamente menor al proceso manual (referencia: 10 horas efectivas).
CA – 02	Porcentaje de registros válidos procesados	El sistema debe procesar el 100% de los registros que contengan serial válido, sin omitir hojas ni archivos.
CA – 03	Comportamiento ante archivos duplicados	Al intentar cargar un archivo ya procesado, el sistema debe identificarlo y omitirlo sin generar errores ni duplicados en la base de datos.
CA – 04	Legibilidad de resultados en la interfaz	Los resultados de búsqueda deben presentarse en formato tabular, mostrando al menos el serial, la referencia, el contrato y el número de AS asociado.
CA – 05	Integridad ante fallos	En caso de error durante la inserción, el sistema debe ejecutar un rollback y no dejar datos parcialmente cargados en la base de datos.

Tabla 2: Criterios de aceptación del sistema

Estos criterios fueron aplicados durante la etapa de validación descrita en la sección 5.4, donde se verificó el cumplimiento de cada uno de ellos mediante pruebas con datos reales del área de Inspecciones.

4. MARCO TEÓRICO

El desarrollo del presente proyecto se fundamenta en conceptos relacionados con la gestión de la información, la calidad de los datos y la automatización de procesos mediante herramientas tecnológicas modernas. Estos elementos permiten transformar información no estructurada en datos organizados y accesibles, facilitando su análisis y uso dentro de entornos organizacionales. A continuación, se presentan los principales conceptos que soportan la solución propuesta.

4.1. Gestión de la información

La gestión de la información se refiere al conjunto de procesos que permiten recolectar, organizar, almacenar y utilizar datos de manera eficiente dentro de una organización. En la actualidad, este concepto ha cobrado mayor relevancia debido al crecimiento exponencial de los datos y a la necesidad de transformarlos en información útil para la toma de decisiones, tendencia ampliamente documentada en el análisis de datos organizacionales (Gartner, 2023).

Una adecuada gestión de la información permite mejorar la eficiencia operativa, reducir la duplicidad de datos y garantizar la disponibilidad de información confiable en tiempo real. En el contexto del área de Inspecciones de Enel Colombia, la gestión de la información representa un reto significativo dado el volumen de archivos AS generados periódicamente y la heterogeneidad de sus estructuras.

4.2. Procesos ETL (Extracción, Transformación y Carga)

Los procesos ETL (Extracción, Transformación y Carga) permiten consolidar información proveniente de múltiples fuentes en una única estructura organizada, facilitando su análisis y consulta. De acuerdo con Amazon Web Services (2022), estos procesos consisten en la extracción de datos desde diferentes orígenes, su transformación para garantizar consistencia y calidad, y su posterior carga en un sistema centralizado. Diversos estudios académicos han analizado los modelos y enfoques de implementación de procesos ETL, destacando su papel central en la consolidación de información proveniente de múltiples fuentes heterogéneas (Vyas y Vaishnav, 2017; Dhaouadi et al., 2022), como ocurre con los archivos AS en el área de Inspecciones, la gestión manual de estos datos puede generar retrasos, reprocesos e inconsistencias. La implementación de un proceso ETL automatizado permite abordar estas limitaciones de manera sistemática y eficiente.

4.3. Calidad de los datos

La calidad de los datos es un factor crítico en los sistemas de información modernos, ya que determina la confiabilidad de los resultados obtenidos. De acuerdo con la norma ISO/IEC

25012 (ISO, 2008), la calidad de los datos se evalúa a partir de características como la exactitud, consistencia, completitud y validez, las cuales permiten garantizar que la información sea adecuada para su uso en procesos de análisis y toma de decisiones.

Diversos estudios destacan que una baja calidad de los datos puede generar reprocesos, errores en la toma de decisiones y pérdidas económicas dentro de las organizaciones (Siebert et al., 2022). En el caso de los archivos AS analizados, se identificaron problemas frecuentes como seriales incompletos, formatos inconsistentes, columnas duplicadas y valores nulos, lo que hace necesario aplicar procesos de limpieza y transformación antes de poder utilizar la información de manera efectiva.

4.4. Bases de datos y centralización de la información

Las bases de datos permiten almacenar información de manera estructurada, facilitando su consulta, actualización y análisis. En particular, los sistemas de gestión de bases de datos relacionales como PostgreSQL permiten gestionar grandes volúmenes de datos garantizando su integridad y consistencia (PostgreSQL Global Development Group, 2024).

En el marco de este proyecto, se utilizó la base de datos PostgreSQL existente en el área de Inspecciones, dentro de la cual se crearon dos nuevas tablas: una destinada al almacenamiento de los registros de materiales procesados y otra para el control de los archivos AS ya cargados al sistema. La centralización de la información en estas tablas contribuye a mejorar la trazabilidad de los datos y a reducir la dispersión de la información en múltiples fuentes.

4.5. Automatización de procesos

La automatización de procesos consiste en el uso de herramientas tecnológicas para ejecutar tareas repetitivas de manera automática, reduciendo la intervención manual. En los últimos años, se ha consolidado como un elemento clave para la eficiencia operativa en las organizaciones, incluyendo el sector energético, donde la automatización de procesos operativos ha demostrado mejoras significativas en la gestión de información y la toma de decisiones (Parra et al., 2024). En el contexto del manejo de datos, la automatización permite agilizar actividades como la limpieza, transformación y consolidación de la información, reduciendo tiempos de procesamiento y errores humanos. En este proyecto, la automatización se implementó mediante un sistema que ejecuta de forma autónoma la lectura de los archivos AS, la detección de encabezados, la limpieza de datos y su carga en la base de datos, eliminando la necesidad de intervención manual en cada uno de estos pasos.

4.6. Herramientas tecnológicas utilizadas

Para el desarrollo de la solución propuesta se seleccionaron herramientas teniendo en cuenta criterios de compatibilidad con la infraestructura existente del área, facilidad de implementación y capacidad de procesamiento requerida.

En primer lugar, se utilizó Python 3.14.3 (Python Software Foundation, 2025) como lenguaje de programación principal, debido a su versatilidad en el manejo de datos y a la amplia disponibilidad de librerías especializadas. Las librerías empleadas fueron las siguientes:

- Pandas: Utilizada para la lectura, manipulación y transformación de los archivos Excel y CSV, permitiendo aplicar los procesos de limpieza y normalización de datos de manera eficiente (McKinney, 2022).
- Psycopg2: Adaptador de PostgreSQL para Python, empleado para establecer la conexión con la base de datos y ejecutar las operaciones de inserción y consulta de registros. La función `extras.execute_values` de esta librería permitió implementar la carga masiva de datos de manera optimizada (Psycopg, 2024).
- Re: Librería estándar de Python utilizada para el procesamiento de expresiones regulares, aplicada en la extracción del número de AS y del contrato a partir del nombre de los archivos.
- Os: librería estándar de Python empleada para la gestión de rutas y archivos del sistema operativo, utilizada en el módulo de carga masiva desde carpeta.

Por otra parte, se empleó PostgreSQL como sistema de gestión de bases de datos, dado que es la plataforma utilizada actualmente por el área de Inspecciones para el almacenamiento de información operativa. La elección de esta herramienta permitió garantizar la integración de la solución con la infraestructura existente de la organización, evitando la implementación de sistemas adicionales.

Finalmente, se utilizó Streamlit (Streamlit Inc., 2024) para la construcción de la interfaz de usuario, ejecutada de manera local dentro del entorno de trabajo del área. Streamlit permite desarrollar aplicaciones web de manera rápida y sencilla utilizando únicamente Python, sin requerir conocimientos avanzados en desarrollo web, lo que facilitó la creación de una herramienta accesible y funcional para los usuarios del área de Inspecciones.

En conjunto, los conceptos abordados en este marco teórico permiten sustentar el desarrollo de la solución, evidenciando cómo la aplicación de procesos ETL, la mejora en la calidad de los datos y la automatización de tareas contribuyen a optimizar la gestión de la información en el área de Inspecciones de Enel Colombia.

5. DESARROLLO

5.1. Etapa 1: Análisis del proceso actual

El desarrollo del presente proyecto inicia con el análisis del proceso actual de gestión y verificación de materiales en el área de Inspecciones de Enel Colombia, con el fin de identificar sus principales limitaciones y oportunidades de mejora.

En el modelo operativo existente, la información relacionada con la entrega de materiales es generada por el área de Logística y publicada en una plataforma interna a través de múltiples archivos en formato Excel. Estos archivos corresponden a las Actas de Salida (AS) asociadas a cada Punto de Despacho de Inventario (PDI) y contienen el detalle de los materiales seriadados entregados a los diferentes contratos, tales como medidores, transformadores de corriente (CT), sellos de seguridad y celdas, entre otros. Cada material registrado en un AS posee un serial único que permite su identificación y trazabilidad dentro del proceso de verificación.

Dias Cierre	ESTADO	AS	SEMAFORO	ESTADO DELIVERY
1	10 FINALIZADO	AS11415		PDFs CARGADOS
0	10 FINALIZADO	AS11416		PDFs CARGADOS
0	10 FINALIZADO	AS11417		PDFs CARGADOS
312	07 REVISAR PDF	AS11418		PDFs CARGADOS
0	10 FINALIZADO	AS11419		PDFs CARGADOS
4	10 FINALIZADO	AS11420		PDFs CARGADOS
-3	10 FINALIZADO	AS11421		PDFs CARGADOS
2	10 FINALIZADO	AS11422		PDFs CARGADOS
306	07 REVISAR PDF	AS11423		PDFs CARGADOS
5	10 FINALIZADO	AS11424		PDFs CARGADOS
2	10 FINALIZADO	AS11425		PDFs CARGADOS

Figura 1: Plataforma interna de logística donde se publican los archivos de despacho de materiales.

Sin embargo, el uso de estos archivos como principal fuente de consulta presenta diversas dificultades. En primer lugar, la información se encuentra altamente fragmentada, ya que cada despacho se registra en un archivo AS independiente, el cual a su vez puede contener múltiples hojas de cálculo con estructuras variables. Esta situación impide contar con una visión consolidada del inventario y dificulta la trazabilidad de los materiales.

Autoguardado 01_DESPACHOS_01_ARCHIVOS_DH20043193-APPLUS NORCONTROL I... Última modificación: Ahora mismo

Archivo Inicio Insertar Dibujar Disposición de página Fórmulas Datos Revisar Vista Automatizar Ayuda Acrobat

Portapapeles Fuente Alineación Número Estilos

Formato condicional Dar formato como tabla Estilos de celda Estilos

Insertar Eliminar Formato Celdas Edición

Sensibilidad Complementos Create a PDF

Comentarios Compartir

G2 =BUSCARV(K2;'https://nealandmassy1td-my.sharepoint.com/personal/operador_plataforma3_massygroup_com/Documents/[Libro1.xlsx]510279'!\$5\$2:\$5\$11084;1;0)

	B	C	D	E	F	G	H	I	J	K
	SERIAL INDIVIDUAL	N° CAJA	UD	UBICACIÓN	FECHA CALIBRACION					
1	0000039033	1	1	B3-04A00	20/02/2025	51027924000000039033	20	00039033		51027924000000039033
2	0000038174	1	1	B3-04A00	20/02/2025	51027924000000038174	20	00038174		51027924000000038174
3	0000039044	1	1	B3-04A00	20/02/2025	51027924000000039044	20	00039044		51027924000000039044
4	0000038180	1	1	B3-04A00	20/02/2025	51027924000000038180	20	00038180		51027924000000038180
5	0000038169	1	1	B3-04A00	20/02/2025	51027924000000038169	20	00038169		51027924000000038169
6	0000038179	1	1	B3-04A00	20/02/2025	51027924000000038179	20	00038179		51027924000000038179
7	0000039043	1	1	B3-04A00	20/02/2025	51027924000000039043	20	00039043		51027924000000039043
8	0000038178	1	1	B3-04A00	20/02/2025	51027924000000038178	20	00038178		51027924000000038178
9	0000038173	1	1	B3-04A00	20/02/2025	51027924000000038173	20	00038173		51027924000000038173
10	0000038176	1	1	B3-04A00	20/02/2025	51027924000000038176	20	00038176		51027924000000038176
11	0000038172	1	1	B3-04A00	20/02/2025	51027924000000038172	20	00038172		51027924000000038172
12	0000039042	1	1	B3-04A00	20/02/2025	51027924000000039042	20	00039042		51027924000000039042
13	0000038171	2	1	B3-04A00	20/02/2025	51027924000000038171	20	00038171		51027924000000038171

SELLO AZUL SELLO AMARILLO 510250 510265 510279 510282 510300

Figura 2: Ejemplo de estructura de archivo AS con fórmula BUSCARV en columna irrelevante.

Autoguardado 01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2... Última modificación: Ayer a las 10:01

Archivo Inicio Insertar Dibujar Disposición de página Fórmulas Datos Revisar Vista Automatizar Ayuda Acrobat

Portapapeles Fuente Alineación Número Estilos

Formato condicional Dar formato como tabla Estilos de celda Estilos

Insertar Eliminar Formato Celdas Edición

Sensibilidad Complementos Copilot Create a PDF

Comentarios Compartir

J15

	A	B	C	D	E	F	G	H	I	J
	CODIGO	SERIAL INDIVIDUAL	N° CAJA	UD	UBICACIÓN	FECHA CALIBRACION	# ESTIBA			
1	510310	51031025000025010069	1	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010069	
2	510310	51031025000025010070	1	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010070	
3	510310	51031025000025010071	1	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010071	
4	510310	51031025000025010072	1	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010072	
5	510310	51031025000025010065	2	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010065	
6	510310	51031025000025010066	2	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010066	
7	510310	51031025000025010067	2	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010067	
8	510310	51031025000025010068	2	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010068	
9	510310	51031025000025010037	3	1	B2-08A00	11/03/2025	ESTIBA 2	20	#N/D	
10	510310	51031025000025010038	3	1	B2-08A00	11/03/2025	ESTIBA 2	20	#N/D	
11	510310	51031025000025010039	3	1	B2-08A00	11/03/2025	ESTIBA 2	20	#N/D	
12	510310	51031025000025010040	3	1	B2-08A00	11/03/2025	ESTIBA 2	20	#N/D	
13	510310	51031025000025010033	4	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010033	
14	510310	51031025000025010034	4	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010034	
15	510310	51031025000025010035	4	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010035	
16	510310	51031025000025010036	4	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010036	
17	510310	51031025000025010009	5	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010009	
18	510310	51031025000025010010	5	1	B2-08A00	11/03/2025	ESTIBA 2	20	51031025000025010010	

510282 510300 510310 531275 510331 280027 CELDAS 510039 SELLO

Figura 3: Ejemplo de estructura de archivo AS con inconsistencias en los datos y columnas no relevantes

A partir de la estructura de los archivos presentada en las figuras anteriores, se realizó un análisis detallado de la información, identificando múltiples problemáticas que afectan directamente la calidad, utilidad y procesamiento de los datos dentro del área de Inspecciones.

En primer lugar, se evidencian columnas que no aportan valor directo al proceso de verificación de materiales, tales como campos relacionados con la ubicación física del material dentro del almacén, número de caja o datos de estiba. Esta información resulta poco relevante para el área de Inspecciones, dado que los materiales ya han sido entregados a los contratos y el proceso se centra principalmente en validar su trazabilidad a partir de los seriales. Asimismo, se identifica la columna de cantidad (UD), la cual en la mayoría de los casos corresponde a un valor constante de uno (1), debido a que cada medidor, transformador u otro activo seriado posee un serial único, por lo que dicho campo no aporta valor significativo para el análisis principal.

No obstante, algunas de estas variables fueron conservadas dentro de la estructura de datos por criterios de integridad de la información y posibles usos futuros, aunque no sean utilizadas directamente en los procesos de consulta y verificación desarrollados en este proyecto.

Adicionalmente, se presentan inconsistencias en la estructura de los archivos, incluyendo la existencia de columnas sin encabezado definido, lo que dificulta su interpretación, así como la duplicidad de información en diferentes columnas. En algunos casos se identifican columnas adicionales que contienen el mismo serial duplicado o transformado mediante fórmulas, sin aportar valor al proceso.

Por otra parte, se identifican archivos que incluyen columnas generadas mediante la función BUSCARV de Excel, que replican información proveniente de otros archivos o fuentes externas. Este tipo de columnas introduce redundancia e inconsistencias, ya que cuando el archivo fuente referenciado no está disponible al momento de la lectura, los valores de estas celdas pueden presentar errores como #N/A o #¡REF!, lo que inutiliza esa información y obliga a descartarla durante la etapa de limpieza.

Asimismo, se evidencian problemas en la calidad de los registros, tales como la presencia de seriales incompletos, valores nulos o formatos inconsistentes, lo que impide realizar una trazabilidad adecuada de los materiales y obliga a implementar procesos adicionales de validación y depuración antes de poder utilizar la información de manera efectiva.

Finalmente, se observa que información clave para el proceso, como el número de AS y el contrato asociado, no se encuentra registrada dentro de la estructura de las tablas de datos, sino únicamente en el nombre del archivo. Si bien la mayoría de los archivos AS siguen un formato estándar definido por el área de Logística, como por ejemplo 01_DESPACHOS_01_ARCHIVOS_DH20043193_APPLUS_NORCONTROL_INSPECCIONES_AS12843_ARCHIVOS_OPERACIÓN_AS12843.LISTADO_SERIALES.113621, existen casos en los que los archivos son enviados directamente por correo electrónico u otras vías alternas, lo que puede generar variaciones en dicho formato. Esta situación requiere

implementar mecanismos de extracción flexibles que permitan obtener esta información de manera confiable independientemente del formato del nombre del archivo.

Como consecuencia de lo anterior, cuando se presenta una solicitud de verificación por parte de un contrato, el equipo de Inspecciones debe ejecutar un proceso manual que consiste en la descarga de múltiples archivos AS, la importación individual de cada hoja de cálculo en Microsoft Access, la selección manual de las columnas relevantes, la adición de las columnas de contrato y AS en cada hoja y la consolidación de la información mediante consultas de anexo. Este procedimiento implica la realización de tareas repetitivas que, para un volumen de 21 archivos AS con más de 15.000 registros, puede tardar entre dos y tres días en completarse, generando una alta carga operativa y aumentando la probabilidad de errores humanos. Esta dificultad había sido enfrentada de manera recurrente tanto por el encargado de materiales y sistemas del área como por los practicantes anteriores, quienes debían invertir una cantidad significativa de tiempo cada vez que necesitaban localizar un serial o verificar la entrega de un material.

Además, la falta de una herramienta centralizada limita la capacidad de respuesta del área, dificultando la verificación oportuna de la información y afectando la toma de decisiones. Esta situación también representa un riesgo en el control de los activos, ya que la imposibilidad de validar de manera ágil la entrega de materiales puede derivar en inconsistencias, pérdidas de inventario o posibles escenarios de fraude.

A partir de este análisis, se identifica la necesidad de implementar una solución que permita mejorar el proceso de gestión de la información, reduciendo la intervención manual, fortaleciendo la calidad de los datos y facilitando su consulta de manera rápida y confiable.

5.2. Etapa 2: Diseño de la solución

A partir del análisis del proceso actual, se definió el diseño de una solución orientada a optimizar la gestión de la información relacionada con la entrega de materiales en el área de Inspecciones de Enel Colombia. Esta solución se planteó con el objetivo de reducir la carga operativa, mejorar la calidad de los datos y facilitar la consulta de la información de manera ágil y confiable.

El diseño propuesto se fundamenta en la estructuración de un flujo de procesamiento de datos que permite transformar archivos AS no estandarizados en información organizada y

centralizada. Para ello, se definió un proceso que integra las etapas de lectura, limpieza, transformación, almacenamiento y consulta de los datos.

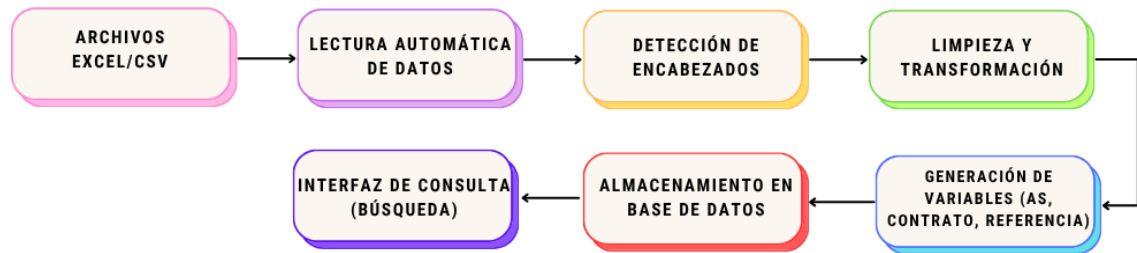


Figura 4: Flujo general del proceso de transformación, almacenamiento y consulta de la información propuesto en la solución

En primer lugar, se contempla la lectura automatizada de los archivos AS provenientes de la plataforma de logística o recibidos por vías alternas, los cuales pueden estar en formato Excel o CSV y contener múltiples hojas con estructuras variables. Dado que estos archivos no presentan una ubicación fija para los encabezados, se incorporó un mecanismo que permite identificar de manera dinámica la fila que contiene los títulos de las columnas, garantizando así la correcta interpretación de la información sin requerir intervención manual.

Posteriormente, se definió una etapa de limpieza y transformación de los datos, en la cual se normalizan los nombres de las columnas, se eliminan valores nulos o inconsistentes y se filtran únicamente los campos relevantes para el área de Inspecciones. En esta fase también se realiza la generación de la variable referencia, obtenida a partir de los últimos ocho caracteres del serial de cada material, así como la extracción del número de AS y del contrato a partir del nombre del archivo.

Una vez procesada la información, el diseño contempla su almacenamiento en la base de datos PostgreSQL existente en el área de Inspecciones, dentro de la cual se crearon dos nuevas tablas: gm_recovery.materiales, destinada al almacenamiento de los registros de materiales procesados, y gm_recovery.archivos_subidos, utilizada para el control de los archivos AS ya cargados al sistema, evitando así duplicidades en la información almacenada.

Con el fin de sustentar las decisiones tecnológicas adoptadas, se presentan a continuación las matrices de selección evaluadas durante la etapa de diseño.

Criterio		Streamlit	Flask	FastAPI
Curva de aprendizaje	de	Alta (solo Python)	Media (requiere HTML/CSS)	Media (requiere conocimientos de APIs REST)
Velocidad de desarrollo	de	Alta	Media	Media

Compatibilidad con pandas y psycpg2	Nativa	Requiere integración manual	Requiere integración manual
Requiere conocimientos en HTML/JS	No	Sí	Sí
Adecuación al contexto del área	Alta (usuarios no técnicos, uso local)	Baja (orientado a despliegue web)	Baja (orientado a servicios backend)
Puntaje total	5/5	2/5	2/5

Matriz 1: Selección de framework para la interfaz de usuario

Streamlit fue seleccionado por permitir el desarrollo de interfaces web funcionales utilizando únicamente Python, sin requerir conocimientos en desarrollo frontend, lo cual se ajustaba tanto al perfil técnico disponible como a la necesidad de contar con una herramienta de uso local accesible para los usuarios del área.

Criterio	PostgreSQL	MySQL	SQLite
Uso previo en el área	Sí	No	Media (requiere conocimientos de APIs REST)
Capacidad de carga masiva	Alta	Alta	Media
Manejo de concurrencia	Alto	Medio	Requiere integración manual
Licenciamiento	Open source	Open source	Open source
Integración con infraestructura existente	Directa (BD ya en uso)	Requiere nueva instalación	Requiere nueva instalación
Puntaje total	5/5	3/5	2/5

Matriz 2: Selección del motor de base de datos

PostgreSQL fue seleccionado dado que es la plataforma de base de datos ya utilizada por el área de Inspecciones, lo que permitió integrar la solución directamente en la infraestructura existente sin necesidad de instalar ni configurar sistemas adicionales, garantizando compatibilidad y continuidad operativa.

Finalmente, se incorporó una interfaz de usuario desarrollada en Streamlit que permite interactuar con la base de datos de forma sencilla, facilitando la búsqueda de información mediante criterios como serial, referencia o código. Esta interfaz fue diseñada para que los usuarios del área de Inspecciones puedan acceder a la información de manera rápida, sin requerir conocimientos técnicos avanzados, y se ejecuta de manera local en el equipo de trabajo del área.

De manera general, la solución propuesta transforma un proceso manual, repetitivo y propenso a errores en un flujo automatizado que garantiza la consistencia de la información y mejora significativamente los tiempos de respuesta en las actividades de verificación.

5.3. Etapa 3: Desarrollo e implementación

En esta etapa se llevó a cabo la construcción e implementación de la solución propuesta. El desarrollo se orientó a la creación de un sistema que permitiera procesar de manera automática los archivos AS, realizar la transformación de los datos y almacenarlos en la base de datos centralizada para su posterior consulta.

La solución fue estructurada en diferentes módulos funcionales, cada uno enfocado en una etapa específica del flujo de procesamiento de la información, incluyendo el procesamiento de datos, la integración con la base de datos, la consulta de información y la carga de archivos.

5.3.1. Etapa 3.1. Arquitectura y diseño del sistema

Con el fin de representar de manera estructurada la solución desarrollada, se presentan a continuación los diagramas que describen el funcionamiento del sistema, la interacción con el usuario y la organización de los datos.

Diagrama de flujo del proceso:

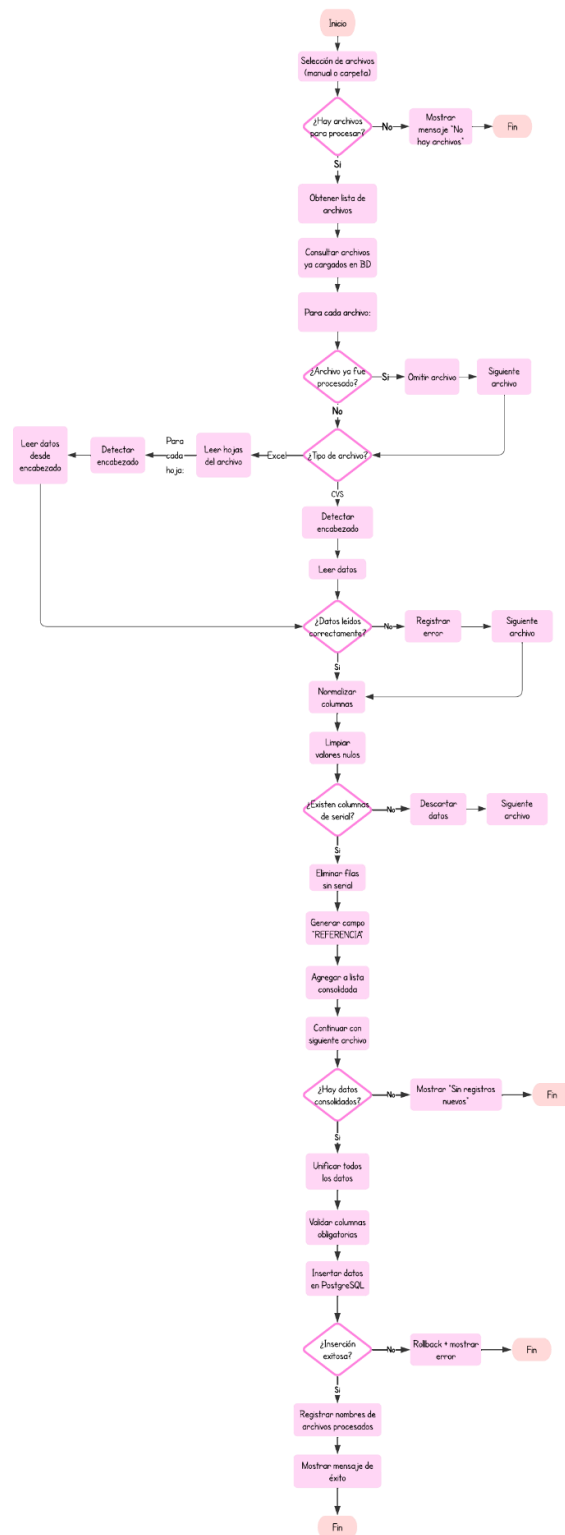


Figura 5: Diagrama de flujo del proceso de carga, procesamiento y almacenamiento de datos

El diagrama de flujo presentado describe el proceso completo de funcionamiento del sistema desarrollado, desde la selección de archivos AS hasta el almacenamiento de la información en la base de datos. El flujo inicia con la selección de los archivos a procesar, ya sea de manera manual o desde una carpeta local. A continuación, el sistema consulta la tabla `gm_recovery.archivos_subidos` para verificar si cada archivo ya fue procesado anteriormente, evitando así duplicidades. Para cada archivo nuevo, se identifican las hojas de cálculo disponibles, se detecta automáticamente la fila de encabezado, se leen los datos y se valida su correcta estructura. Si los datos son válidos, se normalizan las columnas, se eliminan los registros sin serial, se genera el campo de referencia y los datos se agregan a una lista consolidada. Una vez procesados todos los archivos, se unifican los datos, se validan las columnas obligatorias y se realiza la inserción masiva en la tabla `gm_recovery.materiales`. En caso de error durante la inserción, el sistema ejecuta un rollback y muestra un mensaje de error al usuario. Si la inserción es exitosa, se registran los nombres de los archivos procesados en la tabla de control y se muestra un mensaje de confirmación. Este flujo evidencia la automatización completa del proceso, reduciendo la intervención manual al mínimo necesario.

Diagrama de clases UML:

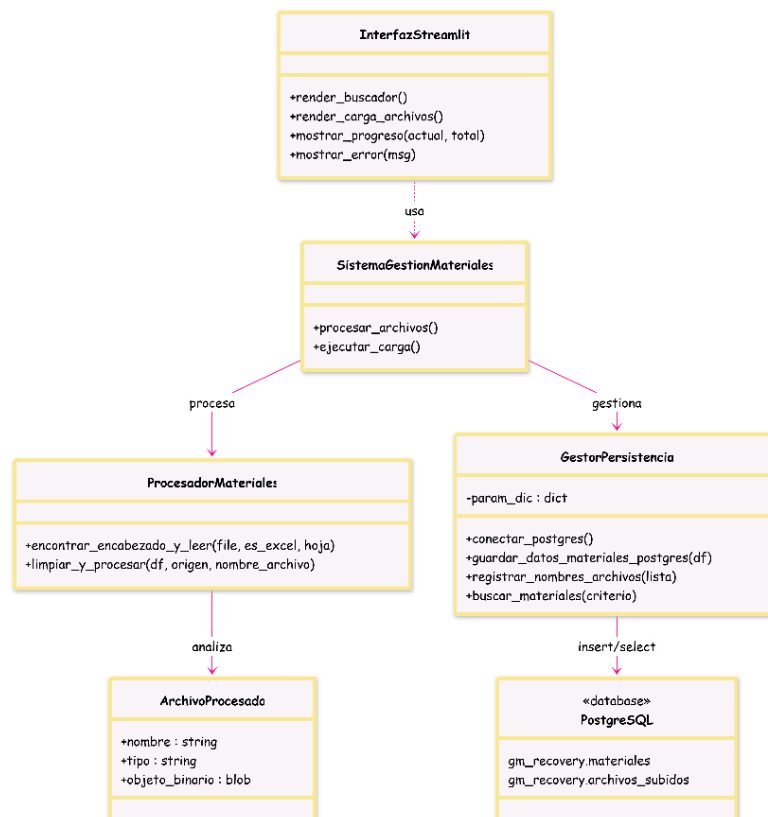


Figura 6: Diagrama de clases del sistema de gestión de materiales

El diagrama de clases describe la estructura interna del sistema, identificando los principales componentes y sus responsabilidades. Se evidencian tres capas principales: la interfaz de usuario, implementada mediante la clase `InterfazStreamlit`, que expone los métodos `render_buscador()`, `render_carga_archivos()`, `mostrar_progreso()` y `mostrar_error()`; el módulo de procesamiento, representado por la clase `ProcesadorMateriales`, responsable de la detección de encabezados y la limpieza y transformación de los datos mediante los métodos `encontrar_encabezado_y_leer()` y `limpiar_y_procesar()`; y el módulo de persistencia, implementado en la clase `GestorPersistencia`, que gestiona la conexión con PostgreSQL y las operaciones de inserción y consulta a través de los métodos `conectar_postgres()`, `guardar_datos_materiales_postgres()`, `registrar_nombres_archivos()` y `buscar_materiales()`. La clase central `SistemaGestionMateriales` orquesta la ejecución del sistema coordinando los módulos anteriores, mientras que la clase `ArchivoProcesado` representa cada archivo AS leído, almacenando su nombre, tipo y contenido en memoria. La base de datos PostgreSQL, representada al final del diagrama, recibe las operaciones de inserción y consulta provenientes del gestor de persistencia sobre las tablas `gm_recovery.materiales` y `gm_recovery.archivos_subidos`.

Esquema de la base de datos:

La tabla `gm_recovery.materiales` almacena el detalle de cada material seriado procesado a partir de los archivos AS. Sus campos son los siguientes: `contrato` (nombre del contrato o PDI al que fue entregado el material), `as` (número del Acta de Salida extraído del nombre del archivo), `codigo` (código del artículo), `serial_individual` (serial único del material, campo principal de trazabilidad), `ud` (cantidad, generalmente igual a 1 por tratarse de materiales seriados), `ubicacion` (ubicación física del material en el momento del despacho), `fecha_calibracion` (fecha de calibración del material), `referencia` (últimos ocho caracteres del serial, generada automáticamente por el sistema), `serie_inicial`, `serie_final` y `serie_chipi` (campos adicionales presentes en algunos tipos de archivos AS), `secuencia` (número de secuencia del registro), `articulo` (descripción del tipo de artículo), `nombre` y `serie` (campos complementarios de identificación del material).

La tabla `gm_recovery.archivos_subidos` lleva el control de los archivos AS que ya han sido procesados e insertados en el sistema. Su único campo es `nombre`, que almacena el nombre del archivo procesado. Esta tabla permite al sistema identificar si un archivo ya fue cargado anteriormente, evitando reprocesos y duplicidades en la información almacenada.

Data Output Messages Notifications											
	CONTRATO character varying	AS. character varying	CODIGO character varying	SERIAL_INDIVIDUAL character varying	UD character varying	UBICACIÓN character varying	FECHA_CALIBRACION character varying	REFERENCIA character varying	SERIE_INICIAL character varying	SERIE_FINAL character varying	SE ch
1	INMEL	AS11587	531275	53127524002409070147	1	B3-01.JP103	2025-04-08 00:00:00	09070147			
2	INMEL	AS11587	531275	53127524002409070146	1	B3-01.JP103	2025-04-08 00:00:00	09070146			
3	INMEL	AS11587	531275	53127524002409070145	1	B3-01.JP103	2025-04-08 00:00:00	09070145			
4	INMEL	AS11587	531275	53127525002510070282	1	B3-01.JP103	2025-04-08 00:00:00	10070282			
5	INMEL	AS11587	531275	53127525002510070281	1	B3-01.JP103	2025-04-08 00:00:00	10070281			
6	INMEL	AS11587	531275	53127525002510070278	1	B3-01.JP103	2025-04-08 00:00:00	10070278			
7	INMEL	AS11587	531275	53127525002510070280	1	B3-01.JP103	2025-04-08 00:00:00	10070280			
8	INMEL	AS11587	531275	53127525002510070279	1	B3-01.JP103	2025-04-08 00:00:00	10070279			
9	INMEL	AS11587	531275	53127525002510070277	1	B3-01.JP103	2025-04-08 00:00:00	10070277			
10	INMEL	AS11587	531275	53127524002409070149	1	B3-01.JP103	2025-04-08 00:00:00	09070149			
11	INMEL	AS11587	531275	53127524002409070148	1	B3-01.JP103	2025-04-08 00:00:00	09070148			

Figura 7: Estructura y contenido de la tabla gm_recovery.materiales en PostgreSQL, donde se almacenan los registros de materiales seriadados procesados a partir de los archivos AS.

5.3.2. Etapa 3.2. Desarrollo del módulo de procesamiento de datos

El módulo de procesamiento de datos fue desarrollado con el objetivo de transformar los archivos AS en formato Excel y CSV con estructuras no uniformes en información organizada y consistente. Para ello, se implementaron funciones que permiten la lectura automática de archivos, la detección dinámica de encabezados y la limpieza de los datos.

Uno de los principales retos abordados fue la variabilidad en la estructura de los archivos AS, ya que estos contienen múltiples hojas y presentan encabezados en diferentes posiciones. Para solucionar esto, se desarrolló una función que analiza las primeras filas de cada archivo en busca de palabras clave predefinidas, permitiendo identificar automáticamente la fila que contiene los nombres de las columnas sin requerir intervención manual.

```
# Palabras que deben estar en el título de cada tabla de los excel
palabras_clave = ['ARTICULO', 'CODIGO', 'SERIAL INDIVIDUAL', 'SERIE', 'UBICACIÓN', 'NOMBRE']

#Recorre fila por fila buscando esas palabras
for i, fila in df.iterrows():
    fila_str = " ".join([str(x).upper() for x in fila.values if pd.notna(x)])
    coincidencias = sum(1 for p in palabras_clave if p in fila_str)
    # Si encuentra más palabras clave en esta fila, asume que es el encabezado
    if coincidencias > max_coincidencias:
        max_coincidencias = coincidencias
        fila_encabezado = i
        if coincidencias >= 2: break
```

Figura 8: Detección automática de encabezados en archivos Excel

El fragmento de código presentado recorre fila por fila el archivo, convierte los valores a texto en mayúsculas y cuenta cuántas palabras clave coinciden en cada fila. La fila con mayor número de coincidencias es identificada como el encabezado. Si se encuentran dos o más

coincidencias en una misma fila, el sistema asume inmediatamente que esa es la fila de encabezado y detiene la búsqueda. De esta manera, el sistema logra adaptarse a diferentes estructuras de archivos AS sin requerir configuración manual para cada uno.

Posteriormente, una vez identificada la fila del encabezado, el archivo es leído nuevamente tomando dicha posición como referencia, estructurando correctamente la información para su posterior procesamiento. Esta lectura se realiza tanto para archivos Excel mediante `pd.read_excel()` como para archivos CSV mediante `pd.read_csv()`, garantizando compatibilidad con los diferentes formatos en los que pueden llegar los archivos AS.

```
# Una vez encontrada la fila, vuelve a leer el archivo completo desde ahí
if es_excel:
    df_final = pd.read_excel(ruta_o_archivo, sheet_name=hoja, skiprows=fila_encabezado, dtype=str)
else:
    df_final = pd.read_csv(ruta_o_archivo, skiprows=fila_encabezado, encoding='latin-1', dtype=str)
```

Figura 9: Lectura del archivo con encabezado identificado

Una vez estructurada la información, se implementó un proceso de limpieza y transformación de los datos. En primer lugar, se estandarizan los nombres de las columnas eliminando espacios en blanco, convirtiendo a mayúsculas y reemplazando espacios internos por guiones bajos, lo que garantiza una estructura uniforme independientemente del formato original del archivo.

```
# Limpia nombres de columnas (quita espacios y pone mayúsculas)
df.columns = [str(c).strip().upper().replace(' ', '_') for c in df.columns]
df = df.loc[:, ~df.columns.duplicated()]
```

Figura 10: Estandarización de columnas

Este proceso permite además eliminar columnas duplicadas que puedan existir en el archivo original, como aquellas generadas mediante fórmulas BUSCARV u otras que repliquen información ya presente en otras columnas.

A continuación, se eliminan los registros que no contienen valores en ninguno de los campos de serial disponibles, garantizando que únicamente se procesen datos válidos para la trazabilidad de los materiales. Esto permite descartar filas vacías, subtotales u otras filas irrelevantes que puedan existir en las hojas de cálculo originales.

```
# Borrar filas que no tengan ningún tipo de serial
cols_presentes = [c for c in cols_check if c in df.columns]
if cols_presentes:
    df = df.dropna(subset=cols_presentes, how='all')
```

Figura 11: Eliminación de registros sin información relevante

Finalmente, se genera automáticamente el campo REFERENCIA a partir de los últimos ocho caracteres del serial del material. Esta variable es clave dentro del sistema, ya que permite realizar búsquedas más ágiles cuando el usuario no dispone del serial completo, facilitando la trazabilidad de los materiales dentro del área de Inspecciones.

```
# REFERENCIA: Saca los últimos 8 caracteres del serial
def obtener_ref(row):
    for c in cols_presentes:
        val = str(row[c])
        if val and val != 'nan' and val != 'None' and len(val) >= 4:
            return val[-8:]
    return ""
```

Figura 12: Generación de la referencia

Como resultado del proceso de transformación y limpieza, se obtiene un conjunto de datos estructurado, consistente y listo para su almacenamiento en la base de datos. Las figuras 13 y 14 muestran ejemplos del resultado final del procesamiento para dos tipos distintos de archivos AS: materiales con serial individual como medidores o transformadores, y materiales tipo sello respectivamente. En ambos casos se puede observar que las columnas están normalizadas, los registros incompletos han sido eliminados y el campo de referencia ha sido generado correctamente, evidenciando la correcta ejecución del proceso ETL implementado.

	A	B	C	D	E	F	G	H	I	J	K
	CONTRATO	AS	CODIGO	SERIAL_INDIVIDUAL	UD	UBICACIÓN	FECHA_CALIBRACION	REFERENCIA	SERIE_INICIAL	SERIE_FINAL	SERIE_CHIPI
1	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	30/10/2024	24120205			
2	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	31/10/2024	24120206			
3	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	01/11/2024	24120207			
4	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	02/11/2024	24120208			
5	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	03/11/2024	24110037			
6	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	04/11/2024	24110038			
7	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	05/11/2024	24110039			
8	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	06/11/2024	24110040			
9	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	07/11/2024	24120225			
10	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	08/11/2024	24120226			
11	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	09/11/2024	24120227			
12	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	10/11/2024	24120228			
13	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	11/11/2024	24120201			
14	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	12/11/2024	24120202			
15	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	13/11/2024	24120203			
16	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	14/11/2024	24120204			
17	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	15/11/2024	24120265			
18	OCA	AS11206	510310	51031024000024100000	1	B2-08PP106	16/11/2024	24120266			

Figura 13: Datos procesados y normalizados tras la etapa de limpieza

	A	B	C	D	E	F	G	H	I	J	K	L	
1	CONTRATO	AS	CODIGO	SERIAL_INDIVIDUAL	UD	UBICACIÓN	FECHA_CALIBRACION	REFERENCIA	SERIE_INICIAL	SERIE_FINAL	SERIE_CHIPI	SECUENCIA	ART
15849	APPLUS	AS12145	531278	53127825002510100000	1	B3-01BP204		10130127					
15850	APPLUS	AS12145	163135					51152001	751152001	751152100	F555Q100/CN0013706		
15851	APPLUS	AS12145	163135					51152101	751152101	751152200	F555Q100/CN0013707		
15852	APPLUS	AS12145	163135					51152201	751152201	751152300	F555Q100/CN0013708		
15853	APPLUS	AS12145	163135					51152301	751152301	751152400	F555Q100/CN0013709		
15854	APPLUS	AS12145	163135					51152401	751152401	751152500	F555Q100/CN0013710		
15855	APPLUS	AS12145	163135					51151501	751151501	751151600	F555Q100/CN0013701		
15856	APPLUS	AS12145	163135					51151601	751151601	751151700	F555Q100/CN0013702		
15857	APPLUS	AS12145	163135					51151701	751151701	751151800	F555Q100/CN0013703		
15858	APPLUS	AS12145	163135					51151801	751151801	751151900	F555Q100/CN0013704		
15859	APPLUS	AS12145	163135					51151901	751151901	751152000	F555Q100/CN0013705		
15860	APPLUS	AS12345	510265	51026524000000100000	1		03/06/2025	105491					
15861	APPLUS	AS12345	510265	51026524000000100000	1		04/06/2025	105492					
15862	APPLUS	AS12345	510265	51026524000000100000	1		05/06/2025	105493					
15863	APPLUS	AS12345	510265	51026524000000100000	1		06/06/2025	105494					
15864	APPLUS	AS12345	510265	51026524000000100000	1		07/06/2025	105495					
15865	APPLUS	AS12345	510265	51026524000000100000	1		08/06/2025	105496					
15866	APPLUS	AS12345	510265	51026524000000100000	1		09/06/2025	105497					

Figura 14: Datos procesados y normalizados tras la etapa de limpieza (sellos)

5.3.3. Etapa 3.3. Integración con la base de datos

Una vez procesados los datos, estos son almacenados en la tabla `gm_recovery.materiales` de la base de datos PostgreSQL utilizada por el área de Inspecciones, con el fin de centralizar la información y facilitar su consulta de manera estructurada.

Para la inserción de los datos se implementó un método de carga masiva utilizando la función `execute_values` de la librería `psycopg2`, la cual permite almacenar grandes volúmenes de registros en una única operación, reduciendo significativamente los tiempos de procesamiento en comparación con inserciones individuales. Este enfoque resulta especialmente útil considerando la cantidad de archivos AS y registros que deben ser procesados, como se evidenció durante la validación con más de 15.000 registros procesados en una sola ejecución.

```
query = f"INSERT INTO gm_recovery.materiales ({columnas}) VALUES %s"
```

```
# 'execute_values': Función para que el insert sea súper rápido
extras.execute_values(cursor, query, valores)
```

Figura 15: Inserción de datos en PostgreSQL

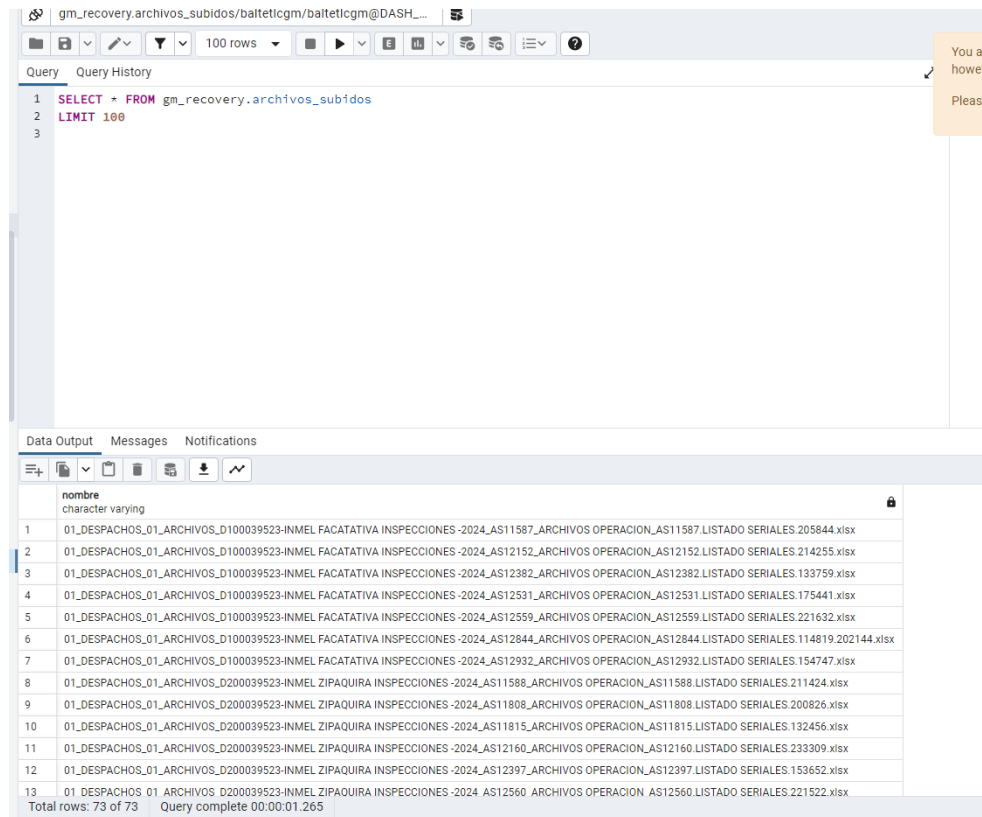
El fragmento de código presentado muestra cómo se construye la consulta de inserción de manera dinámica a partir de las columnas disponibles, y cómo `execute_values` ejecuta la carga masiva de todos los registros en una sola llamada a la base de datos, optimizando el rendimiento del proceso.

Adicionalmente, se implementó un mecanismo de control que evita la duplicidad de información mediante el registro de los nombres de los archivos AS previamente procesados

en la tabla `gm_recovery.archivos_subidos`. Este control permite identificar si un archivo ya ha sido cargado en el sistema antes de procesarlo, evitando reprocesos innecesarios y garantizando la integridad de la información almacenada. La instrucción `ON CONFLICT DO NOTHING` utilizada en la inserción garantiza que, si por alguna razón se intenta registrar un archivo ya existente, la operación no genere errores ni duplicados.

```
# ON CONFLICT DO NOTHING evita errores si se intenta subir el mismo nombre dos veces
cursor.execute("INSERT INTO gm_recovery.archivos_subidos (nombre) VALUES (%s) ON CONFLICT DO NOTHING", (nombre,))
```

Figura 16: Registro y control de archivos procesados



The screenshot shows a PostgreSQL query editor interface. The query entered is:

```
1 SELECT * FROM gm_recovery.archivos_subidos
2 LIMIT 100
3
```

The results are displayed in a table with the following columns and data:

	nombre character varying
1	01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2024_AS11587_ARCHIVOS OPERACION_AS11587.LISTADO SERIALES.205844.xlsx
2	01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2024_AS12152_ARCHIVOS OPERACION_AS12152.LISTADO SERIALES.214255.xlsx
3	01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2024_AS12382_ARCHIVOS OPERACION_AS12382.LISTADO SERIALES.133759.xlsx
4	01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2024_AS12531_ARCHIVOS OPERACION_AS12531.LISTADO SERIALES.175441.xlsx
5	01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2024_AS12559_ARCHIVOS OPERACION_AS12559.LISTADO SERIALES.221632.xlsx
6	01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2024_AS12844_ARCHIVOS OPERACION_AS12844.LISTADO SERIALES.114819.202144.xlsx
7	01_DESPACHOS_01_ARCHIVOS_D100039523-INMEL FACATATIVA INSPECCIONES -2024_AS12932_ARCHIVOS OPERACION_AS12932.LISTADO SERIALES.154747.xlsx
8	01_DESPACHOS_01_ARCHIVOS_D200039523-INMEL ZIPAQUIRA INSPECCIONES -2024_AS11588_ARCHIVOS OPERACION_AS11588.LISTADO SERIALES.211424.xlsx
9	01_DESPACHOS_01_ARCHIVOS_D200039523-INMEL ZIPAQUIRA INSPECCIONES -2024_AS11808_ARCHIVOS OPERACION_AS11808.LISTADO SERIALES.200826.xlsx
10	01_DESPACHOS_01_ARCHIVOS_D200039523-INMEL ZIPAQUIRA INSPECCIONES -2024_AS11815_ARCHIVOS OPERACION_AS11815.LISTADO SERIALES.132456.xlsx
11	01_DESPACHOS_01_ARCHIVOS_D200039523-INMEL ZIPAQUIRA INSPECCIONES -2024_AS12160_ARCHIVOS OPERACION_AS12160.LISTADO SERIALES.233309.xlsx
12	01_DESPACHOS_01_ARCHIVOS_D200039523-INMEL ZIPAQUIRA INSPECCIONES -2024_AS12397_ARCHIVOS OPERACION_AS12397.LISTADO SERIALES.153652.xlsx
13	01_DESPACHOS_01_ARCHIVOS_D200039523-INMEL ZIPAQUIRA INSPECCIONES -2024_AS12560_ARCHIVOS OPERACION_AS12560.LISTADO SERIALES.221522.xlsx

Total rows: 73 of 73 Query complete 00:00:01.265

Figura 17: Tabla archivos procesados en PostgreSQL

Asimismo, aunque durante la etapa de análisis se identificaron columnas que no aportan valor directo al proceso de verificación, algunas de estas fueron conservadas en la base de datos con el fin de mantener la integridad de la información y permitir posibles usos futuros. Sin embargo, dichas variables no son utilizadas en las consultas principales del sistema, las cuales se enfocan en la trazabilidad de los materiales a partir de los seriales, referencias, contrato y número de AS.

5.3.4. Etapa 3.4. Desarrollo de la interfaz de usuario

Con el fin de facilitar la consulta de la información, se desarrolló una interfaz de usuario utilizando Streamlit, ejecutada de manera local en el equipo de trabajo del área. La interfaz cuenta con dos módulos principales: el buscador de materiales y el módulo de carga de archivos, accesibles desde un menú de navegación en la parte superior de la aplicación.

Gestión de Materiales por PDIs

 Buscador  Carga de Archivos

Búsqueda de Material

Ingresa la Serie, Referencia o Código:

Figura 18: Interfaz de búsqueda de materiales

A través del módulo de búsqueda, el usuario puede ingresar un criterio de consulta como el serial completo, la referencia (últimos ocho caracteres del serial) o el código del artículo, ejecutando la consulta directamente sobre la tabla `gm_recovery.materiales` y obteniendo los resultados de manera inmediata. Esto elimina la necesidad de revisar múltiples archivos AS manualmente y permite reducir significativamente los tiempos de respuesta en los procesos de verificación. En caso de que el criterio ingresado no corresponda a ningún registro en la base de datos, el sistema muestra un mensaje informativo al usuario indicando que no se encontraron resultados.

Gestión de Materiales por PDIs

 Buscador  Carga de Archivos

Búsqueda de Material

Ingresa la Serie, Referencia o Código:

	CONTRATO	AS_	CODIGO	SERIAL_INDIVIDUAL	UD	UBICACIÓN	FECHA_CALIBRACION	REFERENCIA	SERIE_INICIAL	SERIE_FINAL	SERIE
0	INMEL	AS11588	510250	51025023000023915412	1	B3-01GP404	2024-07-20 00:00:00	23915412			

Figura 19: Resultado de búsqueda de materiales en la interfaz

Los resultados se presentan en forma de tabla estructurada, permitiendo visualizar de manera clara la información asociada a cada material, incluyendo datos como el serial, la referencia, el contrato y el número de AS. Esto facilita la trazabilidad de los materiales y mejora la toma de decisiones dentro del área de Inspecciones, ya que con una simple búsqueda es posible determinar si un material fue despachado, a qué contrato fue entregado y en qué Acta de Salida quedó registrado.

Con el fin de verificar que la interfaz resultara accesible para los usuarios del área, el encargado de materiales y sistemas utilizó la herramienta durante las pruebas de validación, realizando búsquedas por serial y referencia sobre los registros cargados. Los resultados obtenidos fueron comprensibles para el usuario sin necesidad de instrucciones técnicas adicionales, confirmando que la interfaz cumple con el criterio de usabilidad para personas sin conocimientos avanzados en programación o bases de datos. Esta validación informal permite establecer que la herramienta es funcional no solo desde el punto de vista técnico, sino también desde la perspectiva del usuario final al que está dirigida.

5.3.5. Etapa 3.5. Implementación del módulo de carga

Adicionalmente, se implementó un módulo de carga que permite procesar archivos AS de manera manual o masiva, facilitando la actualización continua de la base de datos a medida que el área de Logística genera nuevos despachos.



Figura 20: Módulo de carga de archivos

Este módulo ofrece dos opciones de carga. La primera permite cargar archivos de manera manual mediante una interfaz de arrastrar y soltar, aceptando archivos en formato XLSX y CSV con un límite de 200 MB por archivo. La segunda opción permite realizar una carga masiva desde una carpeta local del equipo, en la cual el usuario especifica la ruta de la carpeta

que contiene los archivos AS a procesar y el sistema los lee y procesa de manera automática uno por uno.

```
# Agregar archivos manuales si existen
if archivos_manuales:
    for a in archivos_manuales:
        archivos_a_procesar.append({'nombre': a.name, 'obj': a, 'tipo': 'manual'})

# Agregar de X carpeta si existe
if os.path.exists(ruta_input):
    for f in os.listdir(ruta_input):
        if f.endswith(('.xlsx', '.csv')):
            archivos_a_procesar.append({'nombre': f, 'obj': os.path.join(ruta_input, f), 'tipo': 'local'})
```

Figura 21: Proceso de carga masiva

Carga Masiva desde Carpeta

Nombre de la carpeta:

 Iniciar Procesamiento Total

Procesando (84/84): AS11413 APPLUS NORCORCONTROL INSPECCIONES (1).xlsx

 4078 registros insertados con éxito.

¡TERMINADO! Se procesaron 11 archivos y ya están subidossss

Figura 22: Ejecución del proceso de carga masiva de archivos

Durante la ejecución del proceso, el sistema realiza de manera automática la lectura, transformación y almacenamiento de los datos, mostrando mensajes de estado que permiten al usuario conocer el avance del procesamiento, los archivos que ya habían sido procesados anteriormente y la cantidad de registros cargados exitosamente. En caso de producirse algún error durante la inserción, el sistema ejecuta un rollback de la transacción y muestra un mensaje descriptivo del error, garantizando que no queden datos parcialmente insertados en la base de datos. Esto garantiza un proceso eficiente, controlado y sin intervención manual.

5.4. Etapa 4. Validación de la solución

Con el propósito de validar el funcionamiento del sistema desarrollado y medir su impacto en el proceso del área de Inspecciones, se realizaron pruebas utilizando datos reales provenientes del área de Logística, correspondientes a 21 archivos AS del contrato OCA, los cuales contenían en conjunto más de 15.000 registros de materiales seriados.

La validación se enfocó en comparar el proceso manual previamente utilizado frente a la solución implementada, mediante el análisis de indicadores clave de desempeño (KPIs), tales como el tiempo de procesamiento, la carga operativa, la calidad de los datos y la trazabilidad de la información.

5.4.1. Etapa 4.1. Metodología de validación

Para la evaluación de la solución, se tomaron como referencia los mismos 21 archivos AS del contrato OCA, los cuales habían sido procesados previamente de manera manual por la autora utilizando Microsoft Access. Este proceso manual consistía en la importación individual de cada hoja de cálculo, la selección manual de las columnas relevantes, la adición de las columnas de contrato y AS en cada hoja y la consolidación de la información mediante consultas de anexo, repitiendo este procedimiento para cada una de las hojas contenidas en los 21 archivos.

Se realizó una comparación directa entre:

- El proceso manual, basado en la importación y consolidación hoja por hoja en Microsoft Access, con selección manual de columnas y adición de variables clave en cada paso.
- El proceso automatizado, mediante el sistema desarrollado, que realiza la lectura, limpieza, transformación y carga de todos los archivos en una única ejecución.

Los tiempos del proceso manual fueron medidos directamente por la autora durante la ejecución de dicha tarea, con base en su experiencia realizando este proceso en el marco de la práctica profesional. Es importante reconocer que esta condición introduce un sesgo de medición, dado que la misma persona que ejecutó el proceso manual fue también quien desarrolló el sistema automatizado. Sin embargo, esta limitación no invalida la comparación: los tiempos registrados se tomaron como línea base de referencia a partir de la experiencia directa con el proceso real, y son consistentes con lo reportado por el encargado de materiales del área, quien confirmó que dicho proceso representaba una carga operativa significativa y recurrente. Por su parte, los tiempos del sistema automatizado fueron medidos durante la ejecución de la herramienta sobre los mismos 21 archivos AS.

En cuanto al alcance de las pruebas, la validación formal del sistema se realizó con 21 archivos AS correspondientes al contrato OCA, los cuales representaban el conjunto de datos disponible con el que la autora había ejecutado previamente el proceso manual en Microsoft Access, lo que permitió establecer una comparación directa y objetiva. No obstante, durante el desarrollo del proyecto también se procesaron archivos correspondientes a los contratos APPLUS e INMEL, verificando que el mecanismo de detección dinámica de encabezados funcionara correctamente ante los diferentes formatos de archivos AS presentes en el área. En todos los casos evaluados, el sistema logró identificar automáticamente la fila de encabezado y procesar los registros sin requerir configuración manual por archivo. Cabe

señalar que en los casos en que el nombre del archivo no sigue el formato estándar definido por el área de Logística, el campo de contrato y número de AS puede requerir verificación manual, lo cual constituye una limitación conocida del sistema documentada en las conclusiones del presente trabajo.

Durante las pruebas también se evaluó la capacidad del sistema para detectar archivos previamente procesados, verificando que el mecanismo de control implementado en la tabla `gm_recovery.archivos_subidos` funcionara correctamente ante intentos de carga duplicada.

5.4.2. Etapa 4.2. Análisis de indicadores

Tiempo de procesamiento

- Antes: El procesamiento manual de los 21 archivos AS del contrato OCA, con más de 15.000 registros en total, tomó aproximadamente dos días de trabajo, con sesiones de cinco horas continuas cada día, para un total de aproximadamente diez horas efectivas de trabajo.
- Después: El sistema automatizado procesó los mismos 21 archivos AS y sus respectivos registros en cuestión de segundos, realizando de manera automática la lectura de cada hoja, la detección de encabezados, la limpieza de datos, la generación de referencias y la inserción masiva en la base de datos.

Esto representa una reducción superior al 99% en el tiempo requerido para el procesamiento y consolidación de la información.

Carga operativa

- Antes: Alta intervención manual, que implicaba importar cada hoja de cálculo individualmente en Microsoft Access, seleccionar las columnas relevantes, agregar manualmente las columnas de contrato y AS, y ejecutar consultas de anexo para consolidar la información. Este proceso debía repetirse para cada hoja de cada archivo AS, lo que resultaba en una tarea altamente repetitiva y propensa a errores.
- Después: Proceso completamente automatizado que requiere únicamente seleccionar los archivos o especificar la carpeta de origen y ejecutar el sistema. Toda la lectura, transformación y carga se realiza sin intervención manual adicional.

Se evidencia una disminución significativa en la carga operativa del personal, permitiendo enfocar los esfuerzos en actividades de análisis y control en lugar de tareas repetitivas de consolidación de datos.

Calidad de los datos

- Antes: Durante el proceso manual era frecuente la pérdida de registros, ya que al consolidar la información hoja por hoja mediante consultas de anexo existía el

riesgo de omitir hojas, seleccionar columnas incorrectas o no aplicar los mismos criterios de limpieza en todos los archivos, lo que generaba inconsistencias en la base de datos resultante

- Después: El sistema automatizado garantiza que todos los registros válidos de todos los archivos AS sean procesados bajo los mismos criterios de limpieza y transformación, eliminando registros sin serial, normalizando columnas y generando el campo de referencia de manera consistente en cada ejecución. Esto asegura que ningún registro válido sea omitido durante el procesamiento.

Trazabilidad de los materiales

- Antes: La búsqueda de un serial específico requería revisar manualmente múltiples archivos AS o construir primero la base de datos en Access, lo que podía tomar varios días antes de poder responder a una solicitud de verificación.
- Después: Con la información centralizada en la base de datos, la búsqueda de cualquier material por serial, referencia o código se realiza en segundos a través de la interfaz de Streamlit, permitiendo al área de Inspecciones responder de manera inmediata a cualquier solicitud de verificación.

Control de duplicados

Durante las pruebas de validación se verificó que el sistema detecta correctamente los archivos AS que ya habían sido procesados e insertados en la base de datos. Al intentar cargar nuevamente un archivo ya existente, el sistema lo identifica consultando la tabla `gm_recovery.archivos_subidos` y lo omite del proceso de carga, informando al usuario que dicho archivo ya había sido procesado anteriormente. Esto garantiza la integridad de la información almacenada y evita la duplicidad de registros.

Control de pérdidas y posibles fraudes

Uno de los impactos más relevantes de la solución es la mejora en la capacidad de control de los materiales entregados. Antes de la implementación, la dificultad para verificar rápidamente la información limitaba la detección de inconsistencias, lo que podía derivar en pérdidas de inventario o posibles situaciones de fraude. Con la herramienta desarrollada, es posible validar de manera inmediata si un material fue despachado por el área de Logística a un PDI determinado, lo que permite identificar anomalías en tiempo real y fortalecer los mecanismos de control. Este aspecto representa un beneficio económico para la organización, al reducir riesgos asociados a la pérdida de activos.

5.4.3. Etapa 4.3. Resultados de la validación

Los resultados obtenidos evidencian una mejora significativa en la eficiencia del proceso de gestión de la información en el área de Inspecciones. La prueba realizada con los 21 archivos

AS del contrato OCA, correspondientes a más de 15.000 registros de materiales seriados, permitió comparar de manera objetiva y cuantitativa el proceso manual frente al sistema automatizado. La tabla 1 resume los indicadores medidos:

Indicador	Antes (Manual)	Después (Automatizado)	Mejora
Velocidad de procesamiento	~ 600 minutos	< 1 minuto	99,8%
Intervención manual	100%	~ 5%	95%
Calidad y estandarización	Variable	100% consistente	-
Tiempo de búsqueda de serial	2-3 días	< 5 segundos	~ 99,9%
Control de duplicados	0% detectados	100% detectados	-

Tabla 3: Comparación cuantitativa de indicadores antes y después de la implementación

A partir de los datos presentados, se destacan los siguientes resultados:

- El tiempo de procesamiento se redujo de aproximadamente 600 minutos de trabajo efectivo a menos de 1 minuto de ejecución automatizada, representando una mejora del 99,8% en este indicador.
- La intervención manual se redujo del 100% al 5%, limitándose únicamente a la selección de archivos y ejecución del sistema, eliminando todas las tareas repetitivas de importación, selección de columnas y consolidación de datos.
- La calidad y estandarización de los datos pasó de ser variable e inconsistente a garantizar el procesamiento del 100% de los registros válidos bajo los mismos criterios en cada ejecución, eliminando el riesgo de omisión presente en el proceso manual.
- El tiempo de búsqueda de un serial específico se redujo de varios días de trabajo previo a menos de 5 segundos a través de la interfaz de consulta, mejorando en un 99,9% la capacidad de respuesta del área ante solicitudes de verificación.
- El sistema demostró detectar el 100% de los archivos previamente procesados durante las pruebas de validación, garantizando la integridad de la información almacenada y evitando duplicidades en la base de datos.

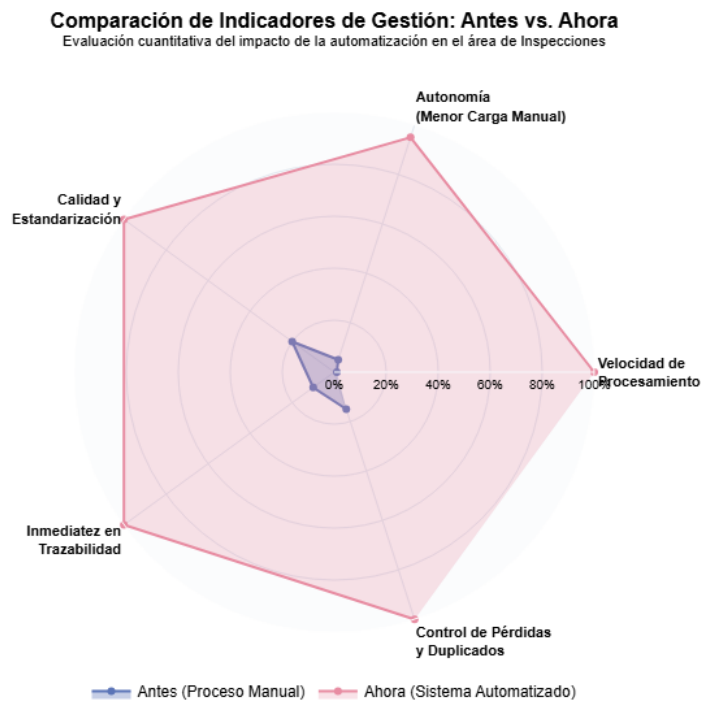


Figura 23: Comparación de indicadores antes y después de la implementación

El gráfico de radar presentado permite visualizar de manera integral el impacto de la automatización en los cinco indicadores evaluados simultáneamente. La diferencia entre el proceso manual, representado por la figura azul concentrada casi en el centro del gráfico, y el sistema automatizado, representado por la figura rosada que abarca la mayor parte del área disponible, evidencia de forma contundente la mejora obtenida en cada una de las dimensiones analizadas: velocidad de procesamiento, autonomía, calidad y estandarización, inmediatez en trazabilidad y control de pérdidas y duplicados.

CONCLUSIONES

A partir del desarrollo e implementación del sistema propuesto, se logró evidenciar una mejora significativa en el proceso de gestión y verificación de la información en el área de Inspecciones de Enel Colombia.

En primer lugar, se cumplió el objetivo principal del proyecto al transformar un proceso manual, basado en la importación y consolidación hoja por hoja de múltiples archivos AS en Microsoft Access, en un sistema automatizado capaz de procesar, limpiar y consolidar la información de manera eficiente. Esto permitió reducir el tiempo de procesamiento de aproximadamente diez horas efectivas de trabajo a tan solo unos segundos para un conjunto de 21 archivos AS con más de 15.000 registros, lo que impacta directamente en la productividad del área y en la capacidad de respuesta ante solicitudes de verificación.

En segundo lugar, la implementación del proceso de limpieza y transformación de datos permitió mejorar la calidad de la información consolidada, garantizando que todos los registros válidos fueran procesados bajo los mismos criterios de normalización, eliminando inconsistencias, valores nulos y registros irrelevantes. A diferencia del proceso manual, en el cual existía el riesgo de omitir hojas, seleccionar columnas incorrectas o aplicar criterios distintos en cada archivo, el sistema automatizado garantiza uniformidad en cada ejecución, lo que se traduce en una mayor confiabilidad de los datos utilizados para la toma de decisiones.

En tercer lugar, la integración de la información en la base de datos PostgreSQL existente del área facilitó la trazabilidad de los materiales, permitiendo realizar consultas rápidas y precisas mediante criterios como serial, referencia o código a través de la interfaz desarrollada en Streamlit. Esta capacidad representa una mejora sustancial frente al modelo anterior, en el cual localizar un serial específico podía requerir varios días de trabajo previo antes de poder dar una respuesta al contrato solicitante.

En cuarto lugar, el desarrollo de la interfaz de usuario permitió que el sistema fuera accesible para los usuarios del área sin requerir conocimientos técnicos avanzados, lo que facilita su adopción dentro del entorno operativo de Inspecciones. La interfaz permite tanto la búsqueda de materiales como la carga de nuevos archivos AS de manera sencilla e intuitiva, integrando en una sola herramienta todas las funcionalidades necesarias para el proceso de verificación.

Finalmente, uno de los aportes más relevantes del proyecto es el fortalecimiento en el control de los activos, ya que la solución permite verificar de manera inmediata si un material fue despachado por el área de Logística a un PDI determinado. Esto no solo mejora la capacidad de respuesta del área de Inspecciones ante reportes de no recepción de materiales, sino que

también contribuye a la detección oportuna de inconsistencias, reduciendo riesgos asociados a pérdidas de inventario o posibles fraudes, lo cual tiene un impacto directo en la eficiencia operativa y en los costos de la organización.

Como parte de la madurez técnica del proyecto, es importante reconocer las limitaciones identificadas y las oportunidades de mejora que quedan abiertas para trabajo futuro. En su estado actual, el sistema opera de manera local sin un servidor dedicado, lo que implica que su disponibilidad depende del equipo en el que se ejecuta. Adicionalmente, el mecanismo de extracción del número de AS y el contrato asociado depende parcialmente del formato del nombre del archivo, por lo que archivos recibidos por vías alternas con nombres no estandarizados pueden requerir ajuste manual en ese campo. El sistema tampoco cuenta actualmente con un módulo de reportes o exportación de resultados, lo cual podría ser de utilidad para el área en procesos de auditoría o seguimiento. Finalmente, las pruebas de validación se realizaron principalmente con archivos del contrato OCA, por lo que la generalización del comportamiento del sistema ante todos los formatos posibles de archivos AS de otros PDIs podría fortalecerse con pruebas adicionales. Estas limitaciones representan oportunidades concretas de mejora para una siguiente versión del sistema, que podría contemplar su despliegue en un servidor interno del área, la incorporación de un módulo de exportación de resultados y la ampliación de las pruebas a la totalidad de los contratos activos.

REFERENCIAS

- Amazon Web Services. (2022). ¿Qué es ETL (extracción, transformación y carga)? <https://aws.amazon.com/es/what-is/etl/>
- Dhaouadi, A., Bousselmi, K., Gammoudi, M. M., Monnet, S., y Hammoudi, S. (2022). Data warehousing process modeling from classical approaches to new trends: Main features and comparisons. *Data*, 7(8), 113. <https://doi.org/10.3390/data7080113>
- Gartner. (2023). Data and analytics trends. <https://www.gartner.com/en/data-analytics/trends>
- International Organization for Standardization. (2008). ISO/IEC 25012: Data quality model. <https://www.iso.org/standard/35736.html>
- McKinney, W. (2022). pandas documentation. https://pandas.pydata.org/docs/getting_started/index.html#intro-to-pandas
- Parra, E., Espinoza, D., y Inga, E. (2024). A step towards digital transformation within a utility company through process automation for a sustainable energy transition. *The Journal of Engineering*, 2024(12). <https://doi.org/10.1049/tje2.70036>
- PostgreSQL Global Development Group. (2024). PostgreSQL documentation. <https://www.postgresql.org/files/documentation/pdf/14/postgresql-14-A4.pdf>
- Psycopg. (2024). Psycopg 2.9 documentation. <https://www.psycopg.org/docs/>
- Python Software Foundation. (2025). What's new in Python 3.14. <https://docs.python.org/3/whatsnew/3.14.html>
- Siebert, J., Joeckel, L., Heidrich, J., Nakamichi, K., Ohashi, K., Namba, I., Knobloch, R., y Trendowicz, A. (2022). Construction of a quality model for machine learning systems. *Software Quality Journal*, 30, 307-335. <https://doi.org/10.1007/s11219-021-09557-y>
- Streamlit Inc. (2024). Streamlit documentation. <https://docs.streamlit.io/get-started/fundamentals>
- Vyas, S., y Vaishnav, P. (2017). A comparative study of various ETL process and their testing techniques in data warehouse. *Journal of Statistics and Management Systems*, 20(4), 753-763. <https://doi.org/10.1080/09720510.2017.1395194>