

Documento de Requerimientos del Proyecto Migrador de Datos (ETL CSV a OLTP)

1. Introducción

1.1. Propósito del Documento

Este documento define los **requerimientos funcionales y no funcionales** del proyecto de migración de datos, cuyo objetivo es implementar un proceso ETL (Extracción, Transformación y Carga) para importar información desde archivos planos en formato CSV hacia una base de datos de un sistema OLTP.

1.2. Alcance

El proyecto migrará datos contenidos en archivos CSV estructurados a una base de datos OLTP, asegurando la integridad, calidad y trazabilidad de los datos importados.

2. Requerimientos Funcionales

2.1. Extracción (Extract)

- El sistema debe permitir la selección manual o automática de archivos CSV desde una carpeta definida.
- El sistema debe validar que el archivo esté correctamente estructurado (cabeceras, separadores, codificación).
- El sistema debe registrar fecha y hora de cada proceso de extracción.

2.2. Transformación (Transform)

- Los datos deben limpiarse y estandarizarse antes de la carga (espacios en blanco, tipos de datos, campos vacíos, etc.).
- Deben implementarse reglas de validación (por ejemplo: campos obligatorios, rangos permitidos, formatos válidos).
- Se deben mapear los campos del archivo CSV con los campos correspondientes de la base de datos destino.

2.3. Carga (Load)

- El sistema debe insertar los datos transformados en la base de datos OLTP.
- El proceso debe realizarse en transacciones para asegurar consistencia (rollback ante fallos).
- Debe existir un mecanismo de control para evitar duplicación de datos.
- El sistema debe generar logs detallados de los datos cargados y los errores encontrados.

2.4. Registro y Auditoría

- El sistema debe registrar quién ejecutó el proceso, cuándo, qué archivo fue procesado y el resultado de este.
- Se debe generar un reporte al finalizar cada carga (registros exitosos, fallidos, causas de error).
- Notificación al finalizar la ejecución vía correo electrónico.

3. Requerimientos No Funcionales

- **Rendimiento:** El proceso de carga debe completarse en menos de 5 minutos para archivos de hasta 1.000.000 registros.
- **Seguridad:** Solo usuarios autorizados podrán ejecutar el proceso ETL y encriptación de contraseñas a conexiones de base de datos y servicios de correos.
- **Escalabilidad:** El sistema debe poder adaptarse para manejar archivos de mayor tamaño en el futuro.
- **Portabilidad:** El sistema debe poder ejecutarse en diferentes entornos (desarrollo, pruebas, producción).
- **Mantenibilidad:** El código debe estar documentado y estructurado para permitir modificaciones futuras.

4. Restricciones

- El archivo fuente estará en formato .csv con separador de coma y codificación UTF-8.
- La base de datos destino es relacional (PostgreSQL).
- La ejecución del ETL puede ser manual (por consola o interfaz) o programada (cron job o similar).

5. Supuestos

- Los archivos CSV serán generados por un sistema externo y entregados en una ubicación compartida.
- Se cuenta con acceso a la base de datos destino para pruebas y producción.
- El modelo de datos del sistema OLTP está definido y documentado.

6. Entregables del Proyecto

- Documento de diseño del proceso ETL.
- Código fuente del proceso ETL.
- Manual de usuario y de operación.

- Informe de pruebas y validaciones.
- Registro de ejecución y errores del proceso.