

# TALLER “ETL”

Docente: Jesus Ariel Bonilla

Estudiante: Paula Andrea Terrios Ossa

|||**TECHSHOP**|||

Corporacion Universitaria del Huila (CORHUILA)

Ingenieria Ambiental

Materia: C.D

Neiva –Huila

2025

# Documentación del Flujo ETL – TECHSHOP

## Introducción

En la actualidad, las empresas generan grandes volúmenes de información provenientes de diferentes fuentes y en distintos formatos. Para que estos datos puedan ser útiles en la toma de decisiones, es necesario aplicar procesos de integración que permitan limpiarlos, transformarlos y almacenarlos en un modelo unificado. El proceso ETL (Extracción, Transformación y Carga) es una de las metodologías más utilizadas para garantizar la calidad de los datos y facilitar su análisis posterior.

## Objetivo

Diseñar y ejecutar un flujo ETL sencillo para la empresa ficticia ***TechShop***, utilizando como fuente un archivo JSON que contiene información de clientes, productos y ventas, con el fin de consolidar los datos en un único repositorio que permita análisis confiables.

## Escenario

***TechShop*** es una tienda en línea de productos tecnológicos que maneja información de sus clientes, catálogo de productos y transacciones de ventas. La información se encuentra en un archivo JSON estructurado que agrupa estos tres conjuntos de datos. El objetivo es integrar dicha información en una tabla final limpia y estandarizada para generar reportes de ventas y conocer el comportamiento de los clientes.

## Descripción de TechShop

***TechShop*** es una empresa dedicada a la venta de productos tecnológicos como computadores portátiles, periféricos y accesorios. Su modelo de negocio se basa en el comercio electrónico y la atención directa a clientes en línea. La empresa busca optimizar la gestión de sus datos para mejorar la toma de decisiones estratégicas, identificar patrones de compra y garantizar que la información de clientes y ventas sea consistente y confiable.

# Procesos ETL

## 1. Extracción

- Fuente: archivo techshop.json.
- Herramienta: pandas en Python.
- Datos extraídos: clientes, productos y ventas.

## 2. Transformación

- Eliminación de valores nulos y duplicados en clientes.
- Normalización de nombres de clientes a mayúsculas.
- Conversión de fechas en ventas a formato estándar ISO (YYYY-MM-DD).
- Validación de cantidades y precios.
- Cálculo de la columna monto = cantidad \* precio.

## 3. Carga

- Base de datos de destino: SQLite.
- Tabla creada: ventas\_limpias.
- Contenido: datos integrados de clientes, productos y ventas, listos para análisis.

## Conclusión

La implementación del flujo ETL permitió transformar datos crudos en información clara, estandarizada y lista para análisis. Se superaron retos como la inconsistencia en formatos de fechas y la presencia de duplicados, lo que demostró la importancia de aplicar reglas de calidad en los procesos de integración. El ejercicio evidenció que incluso trabajando con un único archivo JSON, es necesario un flujo ETL bien definido para garantizar que los datos finales sean útiles y confiables para la toma de decisiones en la empresa *TechShop*.