



Cuatrimestre C3-2025

Asignatura:

Electiva 1

Tema:

ETL Lectura y procesamiento de archivo CSV

Estudiante:

Hideki Sarmiento Ariyama

Matricula:

2024-1453

Profesor:

Francis Ramirez

Fecha:

9/24/2025

1. Un **script SQL** con la creación de la base de datos y sus tablas.

Se encuentra en la entrega

2. El **código del pipeline en Python o C#** que realice la lectura, procesamiento y carga de datos en Github.

Se encuentra en la entrega como zip y en mi Github (<https://github.com/hid-ari/ETL-Lectura-y-procesamiento-de-archivo-CSV>)

3. Una **documentación breve** que explique el flujo de trabajo seguido, acompañada de un diagrama del modelo de datos.

Documentación del Sistema ETL - *Sistema de Análisis de Opiniones de Clientes*

Descripción General

El sistema ETL (Extract, Transform, Load) está diseñado para procesar y cargar datos de diferentes fuentes en la base de datos **MercaJumbo**. El programa procesa archivos CSV de diversas entidades de negocio, aplica reglas de limpieza y validación, y los almacena en la base de datos.

Arquitectura del Sistema

La arquitectura es por capas de la Aplicación

Presentation (Program.cs): Punto de entrada de la aplicación

Business (EtlService.cs): Lógica de negocio y procesamiento ETL

Data: Repositorios para acceso a datos

Entities: Modelos de datos

Componente Principal: EtlService

Responsabilidades

Extracción: Lee archivos CSV utilizando CsvHelper

Transformación: Limpia y valida los datos según reglas de negocio

Carga: Persiste los datos limpios en la base de datos

Entidades Procesadas

1. **Clientes** - Información de usuarios registrados
2. **Productos** - Catálogo de productos con precios y categorías
3. **Fuentes de Datos** - URLs y tipos de fuentes de información
4. **Encuestas** - Respuestas de clientes a preguntas
5. **Comentarios Sociales** - Comentarios en redes sociales con reacciones
6. **Reseñas Web** - Reviews de productos con calificaciones (1-5)
7. **Ventas** - Transacciones de compra
8. **Opiniones** - Calificaciones y comentarios de productos

Reglas de Limpieza de Datos

Validaciones Comunes

Eliminación de espacios en blanco

Validación de campos obligatorios

Detección y eliminación de duplicados

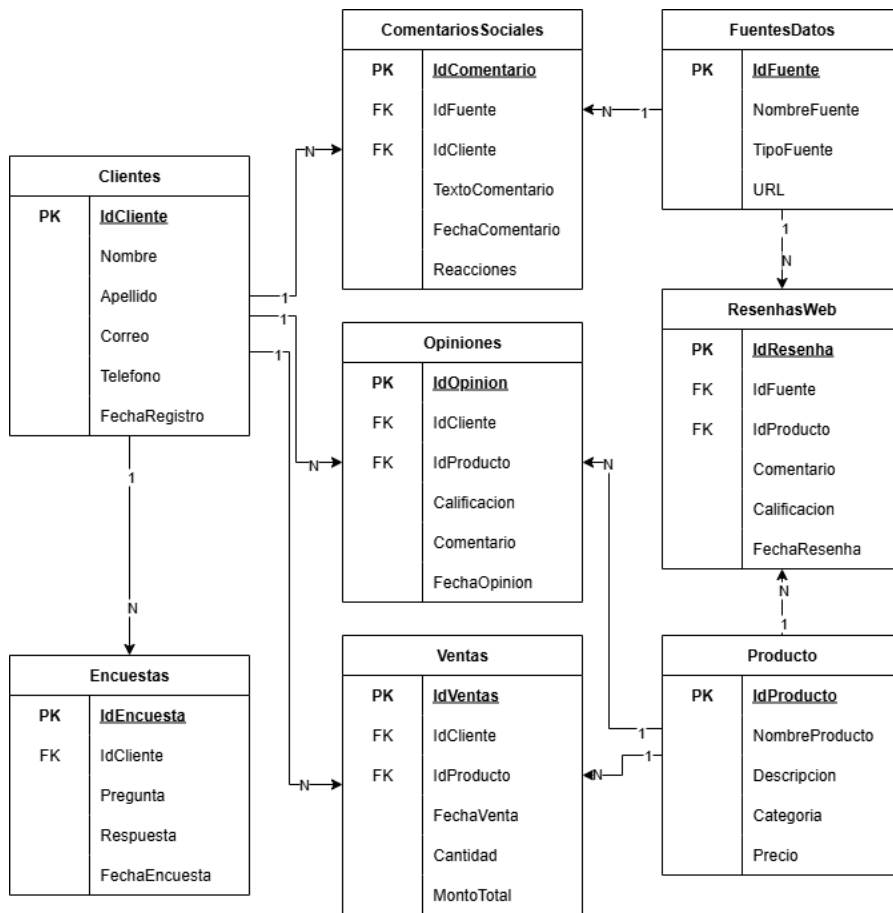
Validaciones Específicas

Clientes: Email único, fecha de registro válida

Productos: Precio mayor a 0, nombre único

Calificaciones: Rango válido de 1-5 estrellas

Ventas: Cantidad y monto mayor a 0



4. Captura de pantallas que muestra la cantidad de registros subida a cada tabla y Result set del select de cada tabla cargada.

Clientes en el CSV

	A	B	C	D	E	F
1	IdCliente	Nombre	Apellido	Correo	Telefono	FechaRegistro
2	1	Juan	Perez	juan.perez@email.com	8095551111	1/15/2024
3	2	Ana	Gomez	ana.gomez@email.com	8295552222	2/10/2024
4	3	Carlos	Ramirez	carlos.ramirez@email.com	8495553333	3/5/2024
5	4	Maria	Lopez	maria.lopez@email.com	8095554444	4/1/2024
6	5	Luis	Martinez	luis.martinez@email.com	8295555555	4/12/2024
7	6	Marcos				
8	4	Maria	Lopez	maria.lopez@email.com	8095554444	4/1/2024

Clientes en la base de datos

	IdCliente	Nombre	Apellido	Correo	Telefono	FechaRegistro
1	1	Juan	Perez	juan.perez@email.com	8095551111	2024-01-15
2	2	Ana	Gomez	ana.gomez@email.com	8295552222	2024-02-10
3	3	Carlos	Ramirez	carlos.ramirez@email.com	8495553333	2024-03-05
4	4	Maria	Lopez	maria.lopez@email.com	8095554444	2024-04-01
5	5	Luis	Martinez	luis.martinez@email.com	8295555555	2024-04-12

Comentarios Sociales en el CSV

	A	B	C	D	E	F
1	IdComentario	IdFuente	IdCliente	TextoComentario	FechaComentario	Reacciones
2	1	1	1	Me encanta comprar en Jumbo	3/5/2024	15
3	2	2	2	El arroz siempre estÃ¡ en oferta	3/6/2024	22
4	3	1	3	La leche subiÃ³ de precio	3/7/2024	10
5	4	4	4	El aceite estÃ¡ mÃ¡s caro este mes	4/3/2024	18
6	5		5	El queso blanco se vende rÃ¡pido		25
7	6	5	5	El queso blanco se vende rÃ¡pido	4/13/2024	26

Comentario Sociales en la base de datos

	IdComentario	IdFuente	IdCliente	TextoComentario	FechaComentario	Reacciones
1	1	1	1	Me encanta comprar en Jumbo	2024-03-05	15
2	2	2	2	El arroz siempre estÃ¡ en oferta	2024-03-06	22
3	3	1	3	La leche subiÃ³ de precio	2024-03-07	10
4	4	4	4	El aceite estÃ¡ mÃ¡s caro este ...	2024-04-03	18
5	5	5	5	El queso blanco se vende rÃ¡...	2024-04-12	25

Encuestas en el CSV

	A	B	C	D	E
1	IdEncuesta	IdCliente	Pregunta	Respuesta	FechaEncuesta
2	1	1	¿Con quÃ© frecuencia compra arroz?	Semanal	3/10/2024
3	2	2	¿Prefiere leche entera o descremada?	Entera	3/11/2024
4	2	2	¿Prefiere leche entera o descremada?	Entera	3/11/2024
5	3			Detergente	
6	3	3	¿QuÃ© productos de limpieza usa mÃ¡s?	Detergente	3/12/2024
7	4	4	¿Compra mÃ¡s en tienda fÃ­sica o en lÃ­nea?	Tienda fÃ­sica	4/2/2024
8	5	5	¿QuÃ© queso prefiere comprar?	Queso blanco	4/12/2024

Encuestas en la base de datos

	IdEncuesta	IdCliente	Pregunta	Respuesta	FechaEncuesta
1	1	1	¿Con quÃ© frecuencia compra arroz?	Semanal	2024-03-10
2	2	2	¿Prefiere leche entera o descremada?	Entera	2024-03-11
3	3	3	¿QuÃ© productos de limpieza usa mÃ¡s?	Deterge...	2024-03-12
4	4	4	¿Compra mÃ¡s en tienda fÃ­sica o en l...	Tienda fÃ...	2024-04-02
5	5	5	¿QuÃ© queso prefiere comprar?	Queso bl...	2024-04-12

FuentesDatos en el CSV

	A	B	C	D
1	IdFuente	NombreFu	TipoFuente	URL
2	1	Facebook	Red Social	https://facebook.com
3	2	Twitter	Red Social	https://twitter.com
4	3	Website O	Web	https://jumbo.com.do
5	4	Instagram	Red Social	https://instagram.com
6	5	YouTube	Video	https://youtube.com
7	6		Web	

FuentesDatos en la base de datos

	IdFuente	NombreFuente	TipoFuente	URL
1	1	Facebook	Red Social	https://facebook.com
2	2	Twitter	Red Social	https://twitter.com
3	3	Website Oficial	Web	https://jumbo.com.do
4	4	Instagram	Red Social	https://instagram.com
5	5	YouTube	Video	https://youtube.com

Opiniones en el csv

	A	B	C	D	E	F
1	IdOpinion	IdCliente	IdProducto	Calificacion	Comentar	FechaOpinion
2	1	1	2	5	Muy buena	3/16/2024
3	2	2	1	4	El arroz es	3/21/2024
4	3	3	3	3	El detergen	3/23/2024
5	4	4	4	5	El aceite e	4/6/2024
6	5	5	5	4	El queso e	4/11/2024
7		4	4	5		

Opiniones en la base de datos

	IdOpinion	IdCliente	IdProducto	Calificacion	Comentario	FechaOpinion
1	1	1	2	5	Muy buena leche, siempre fresca	2024-03-16
2	2	2	1	4	El arroz es de buena calidad, aunque caro	2024-03-21
3	3	3	3	3	El detergente limpia bien pero huele fuerte	2024-03-23
4	4	4	4	5	El aceite es excelente para freir	2024-04-06
5	5	5	5	4	El queso es muy sabroso	2024-04-11

Productos en el CSV

	A	B	C	D	E
1	IdProducto	NombreProducto	Descripcion	Categoria	Precio
2	1	Arroz Premium	Arroz de grano largo de alta calidad	Alimentos	125.5
3	2	Leche Entera	Leche líquida pasteurizada	Lácteos	65.75
4	3	Detergente Azul	Detergente en polvo para ropa	Limpieza	230
5	4	Aceite Vegetal	Aceite de soya refinado		150.25
6	4	Aceite Vegetal	Aceite de soya refinado	Alimentos	150.25
7	5	Queso Blanco	Queso fresco artesanal	Lácteos	300

Productos en la base de datos

	IdProducto	NombreProducto	Descripcion	Categoria	Precio
1	1	Arroz Premium	Arroz de grano largo de alta calidad	Alimentos	125.50
2	2	Leche Entera	Leche líquida pasteurizada	Lácteos	65.75
3	3	Detergente Azul	Detergente en polvo para ropa	Limpieza	230.00
4	4	Aceite Vegetal	Aceite de soya refinado	Alimentos	150.25
5	5	Queso Blanco	Queso fresco artesanal	Lácteos	300.00

ReseñasWeb en el CSV

	A	B	C	D	E	F
1	IdResenha	IdFuente	IdProducto	Calificacion	Comentario	FechaResenha
2	1	3	1	5	Excelente arroz, siempre rinde bastante	3/8/2024
3	2	3	2	4	La leche es buena pero debería estar más barata	3/9/2024
4	3	3	3	3	El detergente limpia bien pero no me gusta el olor	3/10/2024
5	4	3	4	5	El aceite vegetal es de muy buena calidad	4/4/2024
6	4	3	4	5		4/4/2024
7	5	3	5	4	El queso blanco está delicioso	4/11/2024

ReseñasWeb en la base de datos

	IdResenha	IdFuente	IdProducto	Calificacion	Comentario	FechaResenha
1	1	3	1	5	Excelente arroz, siempre rinde bastante	2024-03-08
2	2	3	2	4	La leche es buena pero debería estar más barata	2024-03-09
3	3	3	3	3	El detergente limpia bien pero no me gusta el olor	2024-03-10
4	4	3	4	5	El aceite vegetal es de muy buena calidad	2024-04-04
5	5	3	5	4	El queso blanco está delicioso	2024-04-11

Ventas en el CSV

	A	B	C	D	E	F	
1	IdVenta	IdCliente	IdProducto	FechaVenta	Cantidad	MontoTotal	
2	1	1	2	3/15/2024	3	197.25	
3	2	2	1	3/20/2024	2	251	
4	3	3	3	3/22/2024	1	230	
5	3	3	3	3/22/2024		230	
6	4	4	4	4/5/2024	4	601	
7	5	5	5	4/10/2024	2	600	

Ventas en la base de datos

	IdVenta	IdCliente	IdProducto	FechaVenta	Cantidad	MontoTotal	
1	1	1	2	2024-03-15	3	197.25	
2	2	2	1	2024-03-20	2	251.00	
3	3	3	3	2024-03-22	1	230.00	
4	4	4	4	2024-04-05	4	601.00	
5	5	5	5	2024-04-10	2	600.00	