

**Universidad de San Carlos de Guatemala**

**Facultad de Ingeniería**

**Ingeniería en Ciencias y Sistemas**

**Metodología de Sistemas**

## **Práctica 1**

### **Proceso ETL de la empresa SG-FOOD: extracción, transformación y carga**



**Ponderación: 10 puntos**

**Horas aproximadas de aprendizaje: 16 horas**

## Índice

Resumen Ejecutivo	1
Competencias que desarrollar	1
Objetivos	1
General	1
Específicos	1
Enunciado	2
Descripción del problema a resolver	2
Alcance de la práctica	2
Requerimientos técnicos	3
Entregables	3
Consideraciones	5
Cronograma	5
Rúbrica de calificación	6
Valores	7
Referencias	8

## **Resumen Ejecutivo**

Debido a que los datos se encuentran muchas veces desordenados y mal estructurados las empresas deben transportarlos y cargarlos de manera que en un sistema controlado se puedan adecuar y procesar para poder convertirlos en información crítica y útil, este transporte y carga se le conoce como un proceso ETL (extraer, transformar y cargar). Dependiendo del escenario, contexto y volumen de datos este proceso puede ser corto y dinámico o largo y tedioso, sin embargo, existen herramientas las cuales tienen como propósito el automatizar y acelerar este proceso, una vez logrado la exitosa extracción de sus fuentes y respectivas transformaciones quedan listos para la carga a una herramienta o sistema de análisis multidimensional.

## **Competencias que desarrollar**

Implementa procesos ETL, empleando herramientas especializadas y automatizadas, eficientes asegurando la transformación y carga efectiva de datos para su posterior análisis.

## **Objetivos**

### **General**

- Convertir datos en bruto de múltiples fuentes, utilizando el proceso ETL, a información estructurada para un futuro análisis multidimensional.

### **Específicos**

- Integrar diversas fuentes de datos, aun siendo incompatibles, a un solo proceso de transformación.
- Diseñar un flujo de transformaciones eficiente para la correcta y eficiente estructuración de los datos
- Elaborar un modelo de inteligencia empresarial compatible con el contexto actual de la empresa y los datos provenientes del proceso ETL.

## **Enunciado**

### **Descripción del problema a resolver**

SG-FOOD es una empresa dedicada a la compra, distribución y comercialización de productos alimenticios, manejando diferentes sucursales en diferentes lugares del país. Debido a su constante éxito y crecimiento, ha empezado a enfrentar diversos retos en el manejo de la información tanto de sus ventas y compras, entre ellos:

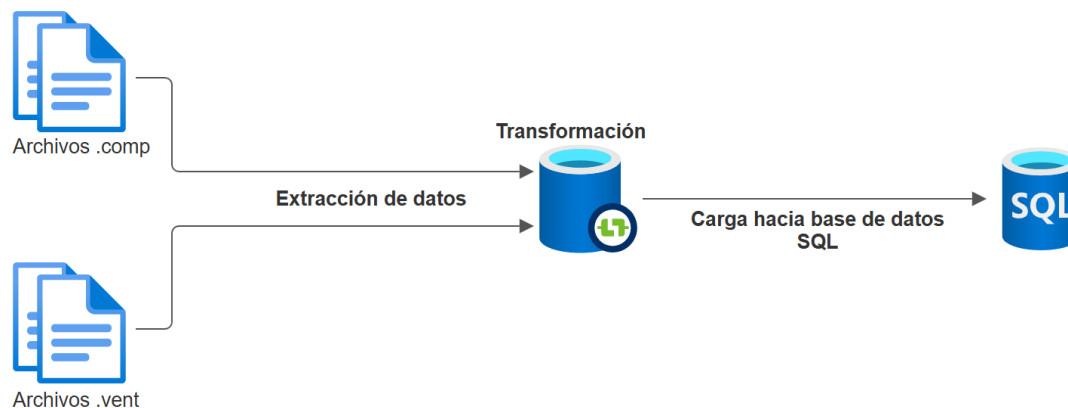
- Diferentes formatos, entre sucursales, en la forma de almacenar sus datos.
- Complejidad al momento de generar reportes generales de la empresa.
- Sobrecarga en las bases de datos transaccionales de las diferentes sucursales.

Esto genera un impacto directo en la toma de decisiones y diseño de estrategias para mejorar y ser una empresa líder en el mercado de productos alimenticios sumado a que en la eficiencia operativa de la empresa ya se notan procesos innecesarios de revisión en grandes volúmenes de datos.

### **Alcance de la práctica**

Se contemplan los siguientes puntos como alcances de la práctica:

- Extraer de manera efectiva los datos de ventas y compras de los últimos meses de la empresa SG-FOOD.
- Homologar de manera efectiva todos los datos provenientes de las diferentes sucursales pertenecientes a SG-FOOD.
- Diseñar el modelo de inteligencia empresarial más adecuado al contexto de SG-FOOD con las operaciones de venta y compras.
- Diseñar consultas enfocadas al análisis, interpretación y comparación de la información cargada en el modelo diseñado.



### Requerimientos técnicos

- Uso de herramientas que permitan la integración de datos con procesos ETL tales como Visual Studio con Microsoft SSIS
- Integración de al menos dos orígenes de datos heterogéneos.
- Modelo de inteligencia empresarial también conocido como modelo multidimensional a elección (modelo estrella, modelo copo de nieve o modelo constelación) almacenado en una base de datos SQL.
- Entorno de pruebas implementando lenguaje SQL para demostrar la carga final y total de los datos.

### Entregables

Todo entregable que se mencione debe estar dentro del repositorio creado para el laboratorio, el repositorio debe respetar el formato de su nombre; “**SS2\_S22025\_#carné**”; agregado cada práctica y proyecto ingresado en el repositorio debe mantener su nombre relacionado en una carpeta, en este caso; “**Practica\_1**”, sin tildes y reemplazando el espacio con un guion bajo.

- Solución y archivos con la implementación completa
- Archivo DDL (.sql) del modelo de inteligencia de negocio creado
- Código o Scripts de limpieza implementados en cualquier lenguaje utilizado

- Consultas de los datos cargados en el modelo en la base SQL:
  - **Total, de compras y ventas por año:** Verifica la distribución temporal de los registros en ambas tablas de hechos.
  - **Productos con pérdida (precio de venta menor al costo de compra):** Identifica productos que generan pérdidas económicas debido a precios de venta más bajos que su costo.
  - **Top 5 productos más vendidos por unidades:** Determina los productos con mayor volumen de ventas, útil para análisis comerciales.
  - **Ingresos por región y año:** Evalúa el rendimiento financiero por región geográfica y temporal.
  - **Proveedores con mayor volumen de compras:** Reconoce a los proveedores estratégicos en función de unidades y costos totales.
- Documentación en formato Mark Down en un archivo README dentro de la carpeta, la documentación debe contener los siguientes puntos:
  - Nombre del proyecto
  - Descripción de cada paso del proceso ETL realizado
  - Explicación de la elección del modelo empresarial implementado
  - Detalle de las características y tablas del modelo empresarial
  - Resultados de consultas y pruebas del modelo en SQL

## Consideraciones

- El modelo empresarial deberá ser diseñado y creado en SQL
- El proceso ETL deberá ser desarrollado por las herramientas indicadas en el laboratorio
- Se deberá agregar como contribuidor al tutor del laboratorio del semestre
- Copias detectadas obtendrán una nota de cero puntos y se reportarán a la Escuela de Ciencias y Sistemas

## Cronograma

Tarea	Fecha
Asignación de la práctica / Entrega del enunciado	30 de julio
Fecha de entrega	15 de agosto
Fecha de calificación	16 de agosto

## Rúbrica de calificación

<b>Criterio</b>	<b>Descripción del criterio a evaluar</b>	<b>Puntos Máximos</b>
<b>Proyecto ETL funcional</b>	La solución (SSIS) extrae correctamente los datos desde ambos orígenes y realiza transformaciones básicas.	20 pts.
<b>Transformación y limpieza</b>	Aplicación correcta de transformaciones como eliminación de nulos, cambios de tipo de datos, normalización u homologación.	15 pts.
<b>Carga a base de datos</b>	Carga exitosa de los datos transformados al modelo empresarial en SQL Server u otro motor compatible.	15 pts.
<b>Modelo empresarial en SQL (DDL)</b>	Definición clara de tablas de hechos y dimensiones con relaciones coherentes. Incluye claves primarias y foráneas.	15 pts.
<b>Diseño de modelo empresarial y justificación</b>	Se entrega un modelo estructurado (estrella, copo de nieve o constelación) y una justificación lógica de su aplicación al caso SG-FOOD.	10 pts.
<b>Consultas realizadas</b>	Evidencia de que se pueden consultar datos de todas las tablas y los datos se encuentran correctamente.	15 pts.
<b>Scripts de limpieza y conteo</b>	Se entregan scripts SQL funcionales para limpieza básica y conteo de registros, acompañados de comentarios.	5 pts.
<b>Documentación técnica en README</b>	README bien redactado con nombre del proyecto, descripción de fases ETL, modelo empresarial y manual básico de implementación.	5 pts.



## Valores

En el desarrollo de la práctica, se espera que cada estudiante demuestre honestidad académica y profesionalismo. Por lo tanto, se establecen los siguientes principios:

- **Originalidad del Trabajo**
  - Cada estudiante o equipo debe desarrollar su propio código y/o documentación, aplicando los conocimientos adquiridos en el curso.
- **Prohibición de Copias y Plagio**
  - Si se detecta la copia total o parcial del código, documentación o cualquier otro entregable, la calificación será de **0 puntos**.
  - Esto incluye la reproducción de código entre compañeros, la reutilización de proyectos de semestres anteriores o el uso de código externo sin la debida referencia.
- **Uso Responsable de Recursos Externos**
  - El uso de bibliotecas, Frameworks y ejemplos de código externos está permitido, siempre y cuando se referencian correctamente y se comprendan plenamente. (Consultar con el catedrático su política)
- **Revisión y Detección de Plagio**
  - Se podrán utilizar herramientas automatizadas y revisiones manuales para identificar similitudes en los proyectos.
  - En caso de sospecha, el estudiante deberá justificar su código y demostrar su desarrollo individual o en equipo. Si este extremo no es comprobable la calificación será de **0 puntos**.
  - Al detectarse estos aspectos se informará al catedrático del curso quien realizará las acciones que considere oportunas.

## Referencias

- Chugugrace. (s. f.). SQL Server Integration Services - SQL Server Integration Services (SSIS). Microsoft Learn. <https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver17>