

**Universidad de San Carlos de Guatemala**  
**Facultad de Ingeniería**  
**Escuela de Ciencias y Sistemas**  
**Seminario de Sistemas 2 Sección A**  
**Ing. Luis Alberto Vettorazzi España**  
**Ing. Fernando Jose Paz Gonzalez**  
**Aux. Jose Fernando Alvarez**  
**Morales**  
**Aux. Sergio Cubur**



## **Proyecto Fase 1**

### **OBJETIVOS:**

- ★ Conocer y aplicar los conocimientos adquiridos en clase sobre DataWarehouse, Datamarts y Business Intelligence.
- ★ Construir una arquitectura para el sistema de BI (Business Intelligence) propuesto.
- ★ Reconocer y aplicar herramientas para un proceso de DataWarehouse dependiendo de las necesidades planteadas por un negocio.
- ★ Conocer y aplicar los conceptos aprendidos en el laboratorio sobre ETL (extracción, transformación y carga de datos).
- ★ Conocer e integrar herramientas de Microsoft para realizar un proceso óptimo de ETL.

### **DESCRIPCIÓN:**

El objetivo principal de este proyecto es que el estudiante pueda proporcionar una solución basada en la implementación de un sistema de análisis y reportes para que la empresa solicitante logre un control sólido sobre sus ventas e inventarios. SG-Food es una megaempresa dedicada a la compra, distribución y comercialización de productos de diversas marcas y categorías. En los últimos años, debido a los cambios en los hábitos de consumo y la creciente tendencia hacia la compra de productos alimentarios y de uso personal para el hogar, las ventas y compras han experimentado un aumento significativo.

En respuesta a este crecimiento reciente, SG-Food necesita optimizar los tiempos de respuesta al realizar análisis sobre sus compras y ventas. Actualmente, el sistema que manejan presenta un tiempo de respuesta demasiado lento y se han enfrentado a problemas en la base de datos principal debido a la cantidad de solicitudes generadas.

Por lo tanto, se le solicita implementar una solución de Business Intelligence que optimice tanto el tiempo de respuesta como la carga sobre la base de datos central de la empresa.

SG-Food requiere una solución que aborde los problemas mencionados, con el objetivo principal de generar reportes sobre las compras y ventas de la empresa. Se busca la capacidad de crear estos informes en intervalos de tiempo flexibles, según las necesidades que se presenten. La información se obtendrá de la base de datos fuente, pero no se podrá operar directamente desde la fuente original, ya que se deben utilizar las herramientas que se explicarán más adelante en el proyecto.

## SG-Food:

Como empresa solicitan una solución que resuelva los problemas anteriormente planteados, esto como fin principal para la generación de reportes sobre las compras y las ventas de la empresa con la facilidad de generar estos en intervalos de tiempo según sean las necesidades que se presenten, la información se obtendrá de la base de datos fuente, pero no se podrá operar desde la fuente original ya que se debe hacer uso de las herramientas explicadas más adelante.

El sistema para la obtención de información tendrá el siguiente funcionamiento:

1. Se le proveerá archivos de texto delimitados por el carácter pipe, pleca o barra vertical (|) con los datos exactos que se encuentran en la base de datos central, **tener en consideración que podrán ser 1 o más archivos de entrada.**

- 1.1. Se contará con dos tipos de archivos para un mejor manejo y comprensión para el análisis de datos con las siguientes extensiones.

|              |  |
|--------------|--|
| <i>.comp</i> | Este contendrá los registros de las compras registradas. |
| <i>.vent</i> | Este contendrá los registros de las ventas registradas.  |

- 1.2. Dichos archivos pueden contener errores en algunos campos como, por ejemplo:

- Campos en blanco.
- Números negativos.
- Números mal escritos.
- Y otros que se consideren necesarios de validar según su criterio.

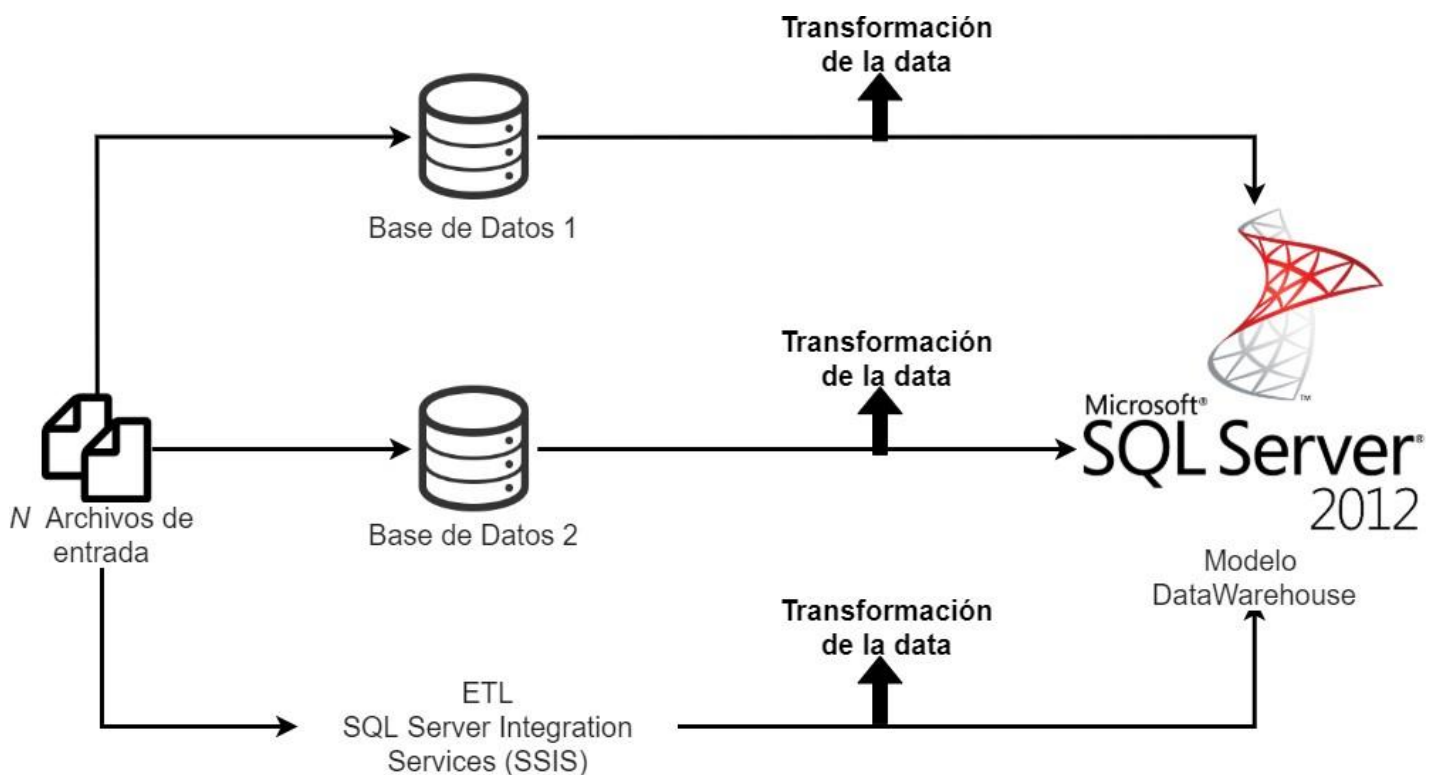
|                 |                 |              |           |
|-----------------|-----------------|--------------|-----------|
| NumeroProveedor | NumeroProveedor | CodProveedor | CostoU    |
| -42022451       | 36khe69         |              | 470.04.18 |

2. En base al **Diagrama 1** se debe conectar 2 bases de datos de distinto fabricante para la extracción de la información contenida en los archivos. Queda a elección del estudiante cuál utilizar de las siguientes opciones:

- MySQL
- Oracle
- Azure
- SQL Server
- O cualquier otra, **exceptuando PostgreSQL.**

3. La carga de los datos se requiere que sea desde 3 orígenes diferentes (2 bases de datos y directo de los archivos) todo esto con las herramientas de business intelligence que ofrece Microsoft Visual Studio.

## FLUJO DEL SISTEMA



## IMPLEMENTACIÓN SUGERIDA

- Cada archivo que se le proporcione tendrá los siguientes nombres:
  - SGFoodXX.comp
  - SGFoodXX.vent

Por lo que se sugiere tener todos estos archivos en una carpeta a la cual tenga acceso total.

- Como primer paso se sugiere la creación de una tabla pivote en cada una de las bases de datos que contenga toda la información proveniente de los archivos de entrada.
- Se deberá crear el modelo para el DataWarehouse esto en SQL Server en el cual se realizarán reportes posteriores.
- Luego de tener la información en los dos orígenes de base de datos relacional seleccionados por el estudiante se deberá realizar el proceso de ETL para luego proceder a la carga de los datos al DataWarehouse.
- Se recomienda iniciar el proceso de ETL con los archivos con la extensión: **.comp** y luego con los archivos con la extensión: **.vent**
- Para un mejor manejo del tiempo a la hora de la calificación se debe utilizar el componente **Execute SQL Task** para vaciar las tablas pivote, así como el modelo del DataWarehouse.
- Por último, se debe crear un script que muestre los totales de elementos insertados en cada tabla de dimensión y de hechos.

## RESTRICCIONES

- No está permitido el uso de PostgreSQL como gestor de base de datos.
- La base de datos para el DataWarehouse debe tener el nombre: ***seminario2\_#carne*** donde ***#carne*** es el número de carné del estudiante y al momento de la calificación está debe estar vacía al igual que las tablas pivotes.
- No se permitirá el uso de tablas intermedias para el manejo de la información, lo permitido serán las tablas pivote y el modelo del DataWarehouse.
- **Es obligatorio** el uso del componente ***Execute SQL Task*** para vaciar las tablas que se utilicen, de no utilizarlo habrá penalización.
- Se debe contar con un script en donde se muestre los totales de registros insertados por cada tabla de dimensión y de hechos.
  - Ejemplo: (*SELECT COUNT(\*) as Totales\_Tabla FROM Tabla*)
- Se evaluará el modelo que fue realizado para la base de datos del DataWarehouse.
- Se evaluará la cantidad de tablas creadas.
- Se evaluará en base a la cantidad de datos ingresados en cada tabla.

## DOCUMENTACIÓN

Para la documentación de este proyecto la empresa solicita que contenga lo siguiente:

1. Nombre del proyecto.
2. Descripción de donde se aplicó cada una de las fases del proceso de ETL.
3. Modelo que se implementó para el DataWarehouse, con su debida justificación.

## CONSIDERACIONES

- La entrega de esta fase es necesaria para tener derecho a calificación a la fase 2.
- Se deberá enviar los archivos necesarios para la calificación: Proyecto de Visual Studio, DDL del DataWarehouse, scripts y documentación.
- Todas las dudas con respecto a esta Fase deberán ser planteadas en los foros creados en la plataforma UEDI.
- Enviar el proyecto vía UEDI en un enlace con su repositorio **el día Lunes 1 de septiembre de 2024 a las 23:59 horas.**
- La calificación será a través de Google Meet en la sala asignada al laboratorio y el estudiante deberá tener la cámara encendida.
- Las entregas tarde se calificarán con un 50% de penalización.
- De encontrar copias se tendrá una nota de 0 y el reporte a la escuela de sistemas.