



Proyecto 1 Fase 1

OBJETIVOS:

- Conocer y aplicar los conocimientos adquiridos sobre DataWarehouse, Business Intelligence.
- Hacer uso de las herramientas adecuadas para la creación de DataWarehouse que se adecue a las necesidades planteadas por el negocio.
- Aplicar los conceptos aprendidos sobre ETL (extracción, transformación y carga de datos).
- Integrar las herramientas de Microsoft para realizar el proceso correcto de ETL.
- Construir la arquitectura del sistema BI (Business Intelligence) según la lógica del negocio.

DESCRIPCIÓN:

El objetivo principal de este proyecto es que el estudiante pueda dar una solución en base a la implementación de un sistema de análisis y reportes para que la empresa solicitante llegue a tener un control sólido sobre sus ventas y de igual forma de sus inventarios.

OneSolution es una cadena de tiendas por departamentos que vende tanto al público en general (por menor) como a clientes mayoristas. Se inició como un negocio pequeño, pero debido al auge que ha tenido la industria alimentaria en esta crisis ocasionada por la pandemia, las ventas y compras han aumentado en gran manera.

Respondiendo a este crecimiento de los últimos meses, la empresa necesita optimizar los tiempos de respuesta al momento de realizar análisis sobre sus ventas y compras, ya que actualmente el sistema que manejan presenta un tiempo de respuesta

demasiado lento y se han presentado problemas en la base de datos principal debido a la cantidad de solicitudes generadas que puede llegar a bloquear las operaciones comerciales, por lo cual se le solicita a usted el implementar una solución de Business Intelligence en la cual el sistema optimice tanto el tiempo de respuesta como el estrés sobre la base de datos central de la empresa.

OneSolution - BI

El proyecto OneSolution-BI, consiste en proveer 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.

Debido a que la empresa busca encontrar una solución que optimice la generación de reportes, desea que se le presenten dos propuestas:

- Utilizando almacenamiento columnar
- Utilizando almacenamiento por filas

Para esto el sistema obtendrá la información de la siguiente manera:

1. Se le proveerá archivos de texto delimitados por el carácter pipe("|") 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>.cp</i>	Este contendrá los registros de las compras registradas.
<i>.vt</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.
2. Deberá cargar los datos utilizando las herramientas de business intelligence que ofrece Microsoft Visual Studio, para la creación de los datawarehouse.

Para lo cual puede realizarlo de dos formas:

2.1. En base a la **Ilustración 1** debe conectarse a una base de datos para la extracción de la información contenida en los archivos. Queda a elección del estudiante cuál utilizar de las siguientes opciones:

- MySQL
- SQLServer
- Oracle
- O cualquier otra, **exceptuando PostgreSQL.**

2.2. En base a la **Ilustración 2** cargando directamente los datos de los archivos para realizar el proceso de ETL.

FLUJO DEL SISTEMA

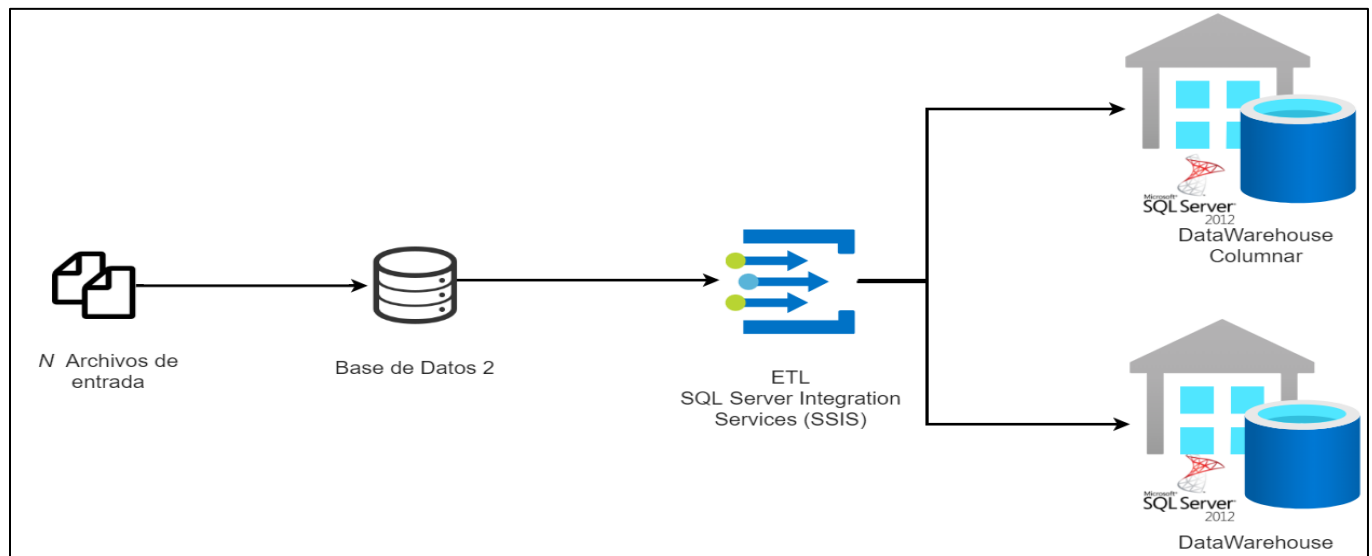


Ilustración 1 Flujo 1 del ETL

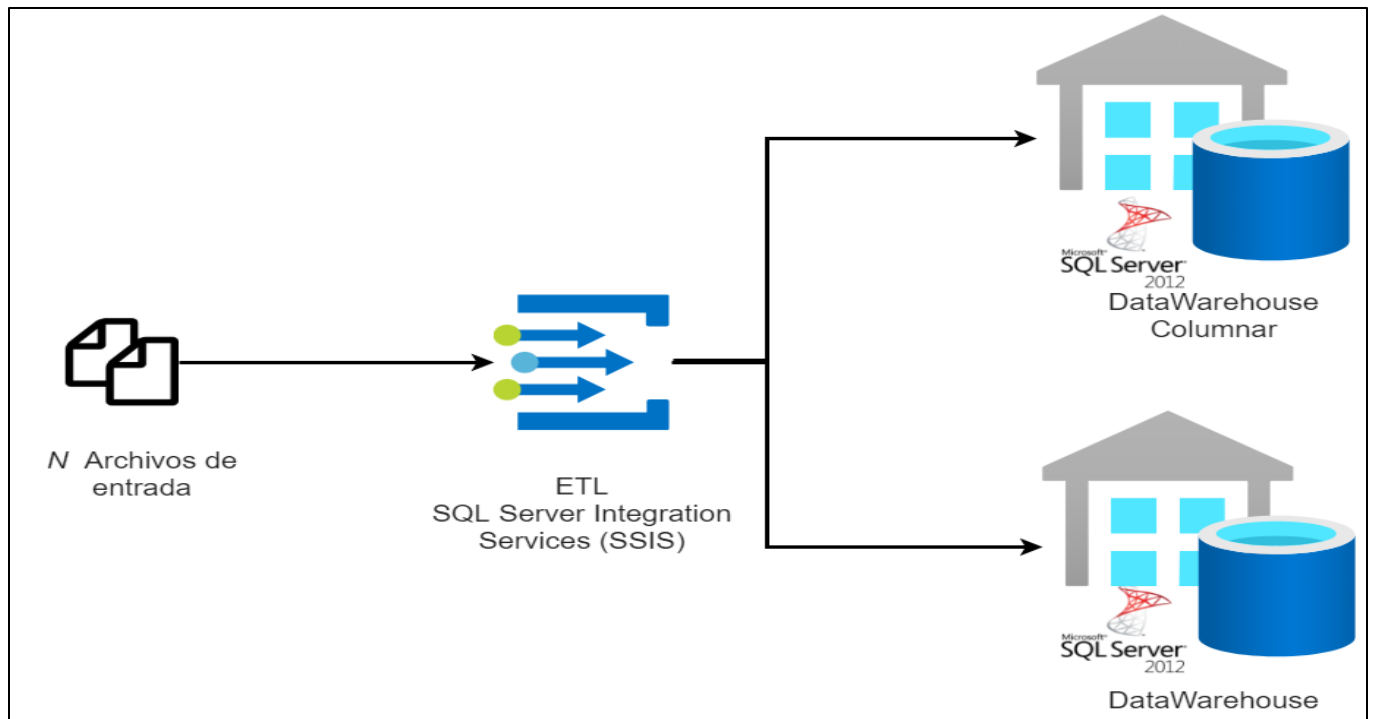


Ilustración 2 Flujo 2 del ETL

IMPLEMENTACIÓN SUGERIDA

Los pasos del 1 al 8 deberán seguirse para cumplir con el Flujo 1 (Ilustración 1), para cargar los datos directamente, Flujo 2, saltarse el paso 2.

1. Cada archivo que se le proporcione tendrá los siguientes nombres:

- OSolutionXX.cp
- OSolutionXX.vt

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

2. 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.
3. Se deberá crear el modelo para el DataWarehouse esto en SQL Server en el cual se realizarán reportes posteriores. Tomar en cuenta que este modelo será usado tanto para almacenamiento columnar como por filas, por lo tanto los siguientes tipos de datos no están soportados:

- Binary, varbinary

- Ntext, text, image
 - varchar (max) y nvarchar(max)
 - uniqueidentifier
 - rowversion, timestamp
 - sql_variant
 - decimal y numeric (con precisión superior a 18 dígitos)
 - Datetimeoffset (con escala mayor que 2)
 - Tipos CLR (hierarchyid y tipos especiales)
 - XML
4. 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. Tomar en cuenta que si está cargando los datos directamente del archivo debe utilizar el componente **Foreach Loop Container**.
 5. Se recomienda iniciar el proceso de ETL con los archivos con la extensión: **.cp** y luego con los archivos con la extensión: **.vt**
 6. 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 como el modelo en cada una de las bases de datos.
 7. Por último se debe crear un script que muestre los totales de elementos insertados en cada tabla de dimensión y de hechos.
 8. Clonar la base de datos del Datawarehouse con los datos ya cargados, crear los índices columnares en para las tablas del nuevo Datawarehouse.

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: **seminariodos#carne** y **semnariodosColumnar#carne** donde **#carne** es el número de carné del estudiante y al momento de la calificación estás deben estar vacías al igual que las tablas pivotes (si las utilizo).
- 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.
- 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 los pasos que se tomaron desde la extracción de los datos hasta la carga al Datawarehouse.
3. Modelo que se implementó para el Datawarehouse, con su debida justificación (Estrella o copo de nieve).

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 o en caso muy especial al correo ss2usac@gmail.com
- Enviar el proyecto vía UEDI en un zip con el nombre: **[SS2]Fase1_carne.zip el día miércoles 9 de Marzo de 2022 a las 23:59 horas.**
- La calificación será vía Google Meet asignado a laboratorio.
- 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.