



DESAFIO 1

Integración de Datos – Consolidado de Facturación de Roaming

Apellidos	Nombres	Carné
Navas de Alvarenga	Sandra Noemy	ND221932

GitHub repositorio: <https://github.com/Sandy-nd/Desaf-o-1>

Introducción

El proyecto tiene como finalidad integrar y consolidar la información de facturación del servicio de roaming de una compañía telefónica que opera en la región centroamericana.

Para ello se proporcionaron tres bases de datos de SQL Server, estas poseen la misma estructura de tablas, que registran la facturación, clientes y tipos de clientes.

Recursos Utilizados

SQL Server 2019

Visual Studio 2022 con SSIS

Flat File Destination (CSV)

GitHub

Resultados

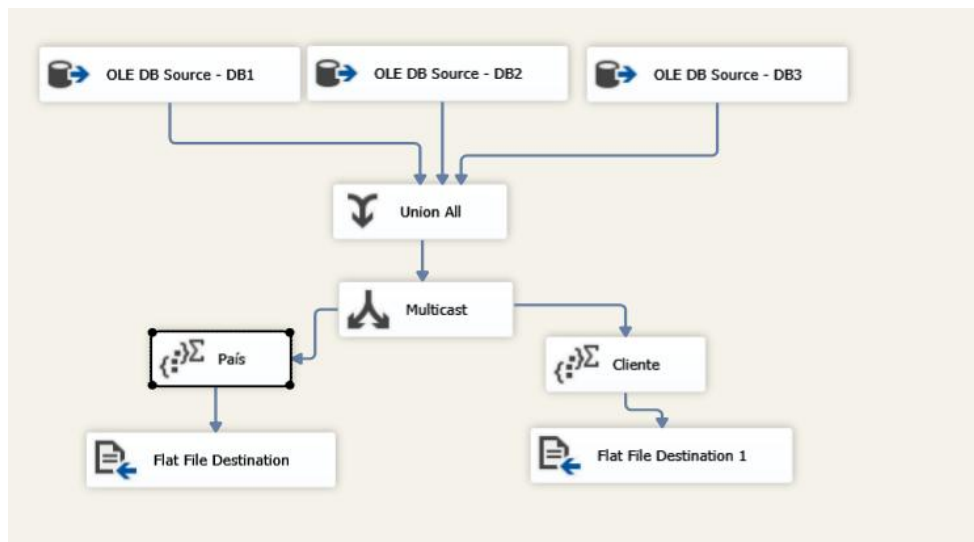
1. Estructura de las bases de datos

Cada base contiene las siguientes tablas principales

- facturacion: Registra las ventas y montos de facturación.
- cliente: Datos de clientes que realizan las compras.
- tipo_clientes: Clasificación de los clientes

2. Proceso ETL

El proyecto se implementó en un único flujo de datos (Data Flow) que cubre las tres etapas del proceso ETL: Extracción, Transformación y Carga.



Datawarehouse y Minería de Datos DMD941 G01T

a) Extracción

Se configuraron tres OLE DB Source para conectarse a cada una de las bases de datos (Costa Rica, Guatemala, Honduras).

En cada origen se utilizó una consulta SQL optimizada que:

- Normaliza el nombre del país.
- Limpia el nombre del tipo de cliente (TRIM, ISNULL, UPPER).
- Convierte los montos a formato numérico.
- Genera un campo de fecha de venta (GETDATE()).

Ejemplo de consulta utilizada (Costa Rica):

The screenshot shows the 'Administrador de conexiones OLE DB' window. The 'OLEDB_Source_DB1' is selected in the dropdown. The 'Modo de acceso a datos' is set to 'Comando SQL'. The 'Texto de comando SQL' field contains the following query:

```
SELECT
    TRIM('COSTA RICA') AS country_name_trim,
    UPPER(TRIM(ISNULL(NULLIF(LTRIM(RTRIM(tc.nombre_tipo)),
    ''), 'UNKNOWN')))) AS customer_type_trim,
    f.idfacturacion AS sale_id,
    GETDATE() AS sale_date,
    CAST(f.monto_facturacion AS NUMERIC(18,2)) AS
    sale_amount_num
FROM facturacion f
INNER JOIN cliente c ON f.idcliente = c.idcliente
INNER JOIN tipo_clientes tc ON c.idtipocliente = tc.idtipocliente
WHERE f.idfacturacion IS NOT NULL;
```

Buttons on the right include: Nueva..., Parámetros..., Generar consulta..., Examinar..., and Analizar consulta.

Datawarehouse y Minería de Datos DMD941 G01T

b) Transformación

Se consolidan y estructuran los datos mediante los siguientes componentes:

Union All: Integra los datos provenientes de las tres bases de datos (Costa Rica, Guatemala y Honduras) en un único flujo de datos unificado.

Union All Transformation Editor			
Configure the properties used to merge multiple inputs into one output by creating mappings between columns.			
ne	Union All Input 1	Union All Input 2	Union All Input 3
	sale_id	sale_id	sale_id
	sale_date	sale_date	sale_date
	country_name_trim	country_name_trim	country_name_trim
	customer_type_trim	customer_type_trim	customer_type_trim
	sale_amount_num	sale_amount_num	sale_amount_num

Multicast: Duplica el flujo consolidado para crear dos rutas de procesamiento distintas, facilitando el análisis multidimensional de los datos sin duplicar el proceso de extracción inicial.

Multicast Output 1	Multicast Output 2
▼ Propiedades comunes	
DeleteOutputOnPathData True	
Description	
ErrorOrTruncationOperatic	
ErrorRowDisposition RD_NotUsed	
ExclusionGroup 0	
HasSideEffects False	
ID 334	
IdentificationString Multicast.Outputs[Multicast Output	
IsErrorOut False	
Name Multicast Output 1	
SynchronousInputID Multicast.Inputs[Multicast Input 1]	
TruncationRowDisposition RD_NotUsed	

Multicast Output 1	Multicast Output 2
▼ Propiedades comunes	
DeleteOutputOnPathData True	
Description	
ErrorOrTruncationOperatic	
ErrorRowDisposition RD_NotUsed	
ExclusionGroup 0	
HasSideEffects False	
ID 347	
IdentificationString Multicast.Outputs[Multicast Output	
IsErrorOut False	
Name Multicast Output 2	
SynchronousInputID Multicast.Inputs[Multicast Input 1]	
TruncationRowDisposition RD_NotUsed	

Datawarehouse y Minería de Datos DMD941 G01T

Aggregate (Consolidado por País): Agrupa la información por país y calcula:

- Conteo de ventas (total_sales) para entender el volumen transaccional por país.
- Sumatoria de montos (total_amount) para analizar el valor económico por región.

Columnas de entrada dispo...

☒ Nombre
☐ (*)
☒ sale_id
☐ sale_date
☒ country_name_trim
☐ customer_type_trim

Input Column	Output Alias	Operation	Compa
country_name_trim	country_name_trim	Group by	
sale_id	total_sales	Count	
sale_amount_num	total_amount	Sum	

Aggregate (Consolidado por Tipo de Cliente): Organiza los datos por categoría de cliente y determina:

- Cantidad de ventas (total_sales) por segmento de cliente.
- Valor total (total_amount) por tipo de cliente, identificando los segmentos más valiosos.

Columnas de entrada dispo...

☒ Nombre
☐ (*)
☒ sale_id
☐ sale_date
☐ country_name_trim
☒ customer_type_trim

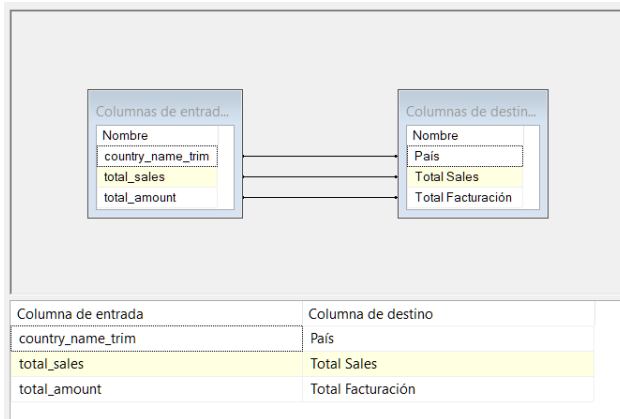
Input Column	Output Alias	Operation	Compa
customer_type_trim	customer_type_trim	Group by	
sale_id	total_sales	Count	
sale_amount_num	total_amount	Sum	

Datawarehouse y Minería de Datos DMD941 G01T

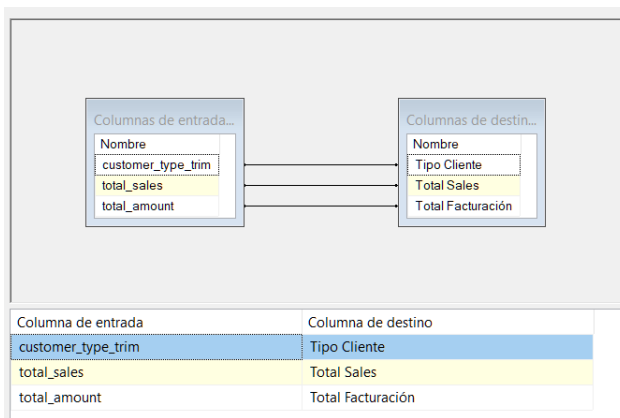
c) Carga

Cada flujo de agregado se conecta a un Flat File Destination que genera un archivo CSV
(Previamente se configuro la configuración de los dos flat files, y la carpeta especifica donde se generan los archivos csv.)

1- Consolidado_Por_Pais.csv



2- Consolidado_Por_TipoCliente.csv



3. Archivos Generados

Nombre	Fecha de modificación	Tipo	Tamaño
Consolidado_Por_Pais.csv	14 sept 2025 04:22 p. m.	Archivo de valores...	1 KB
Consolidado_Por_TipoCliente.csv	14 sept 2025 04:22 p. m.	Archivo de valores...	1 KB

Los archivos pueden abrirse directamente en Excel y se visualizan con las columnas separadas.

- **Consolidado_Por_Pais.csv**

	A	B	C	D
1	País	Total Sales	Total Facturación	
2	COSTA RICA	80	9503.1	
3	HONDURAS	80	8703.1	
4	GUATEMALA	160	16689.8	

- **Consolidado_Por_TipoCliente.csv**

	A	B	C
1	Tipo Cliente	Total Sales	Total Facturación
2	CLIENTE PREFERENCIAL	68	8706.2
3	CLIENTE TURISTA	84	8287.8
4	CLIENTE EJECUTIVO	84	9408.6
5	CLIENTE GUBERNAMENTAL	84	8493.4

Conclusión

Se desarrolló con éxito un proyecto SSIS de integración de datos que consolida información de tres fuentes distintas en un solo flujo, generando reportes limpios y listos para análisis.