

REPOSITORIO: <https://github.com/elopezt/PruebaTecnicaSIMAN>

PARTE 1. MODELADO DE DATOS

Para poder diseñar y crear un modelo de datos relacional adecuado para un conjunto de datos decidí seleccionar las siguientes tablas:

1. **olist_customers_dataset.csv:** Para entender quiénes son los clientes y su ubicación.
2. **olist_order_items_dataset.csv:** Para obtener detalles sobre los productos en cada pedido.
3. **olist_orders_dataset.csv:** Para obtener información general sobre los pedidos, como fechas y estados.

La razón de mi selección se basa en que estas tres tablas permiten modelar las relaciones entre clientes, pedidos y productos, y al ser de las tablas más relevantes proporcionan una base sólida para realizar un análisis más detallado de aspectos importantes generales sobre los pedidos, los ítems en los pedidos y sus clientes.

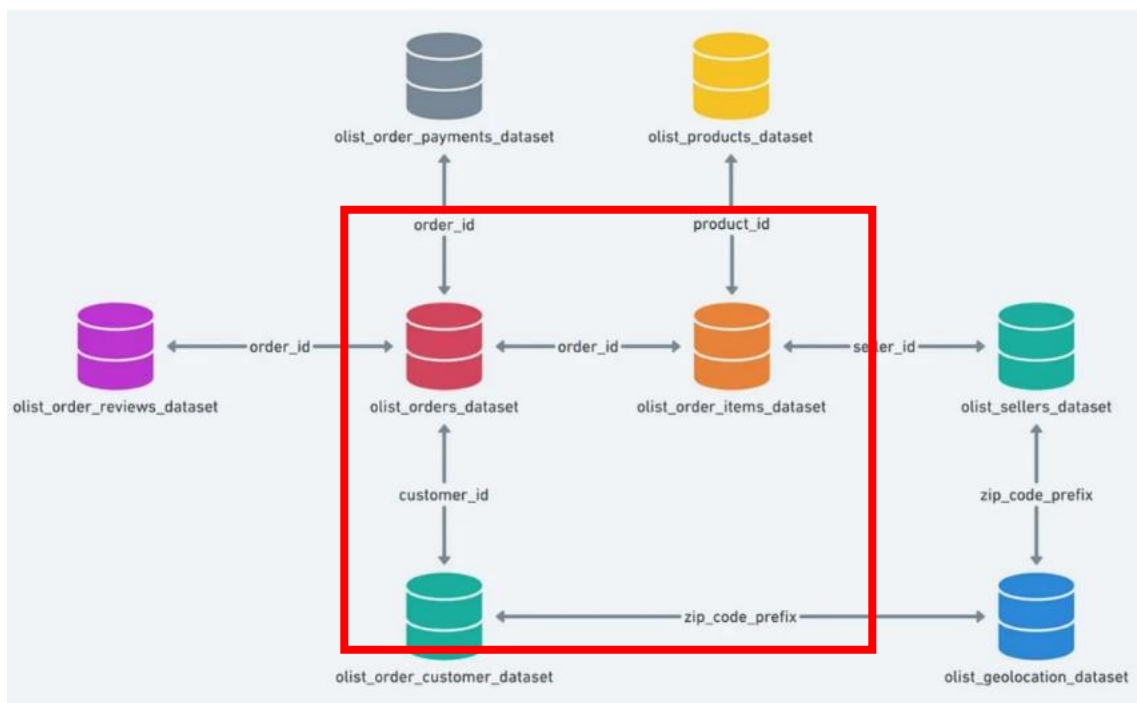


Figura 1. Relaciones entre tablas

El patrón de diseño que definí para el modelo de datos es el de un modelo entidad-relación (ER), ya que este diseño de bases de datos ofrece una forma estructurada y eficiente de diseñar bases de datos relacionales, proporcionando claridad, abstracción y facilidad de mantenimiento en el desarrollo de

sistemas de información. El modelo entidad-relación proporciona una representación visual clara de la estructura de los datos y las relaciones entre las entidades lo cual facilita la comprensión tanto para los diseñadores de bases de datos como para los usuarios finales, permite remover la complejidad de la base de datos al presentarla de una manera más estructurada y fácil de entender, entre otras.

A continuación, se muestra el diseño en base a las 3 tablas seleccionadas:

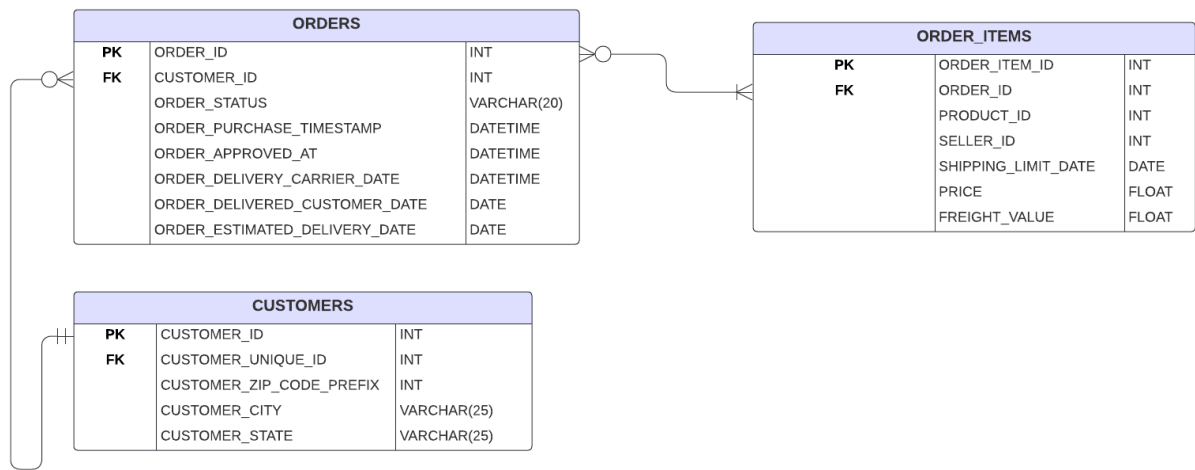


Figura 2. Modelo Entidad Relación

En base a las entidades y atributos de las tablas proseguí con la creación de las tablas haciendo uso de SQL Server.

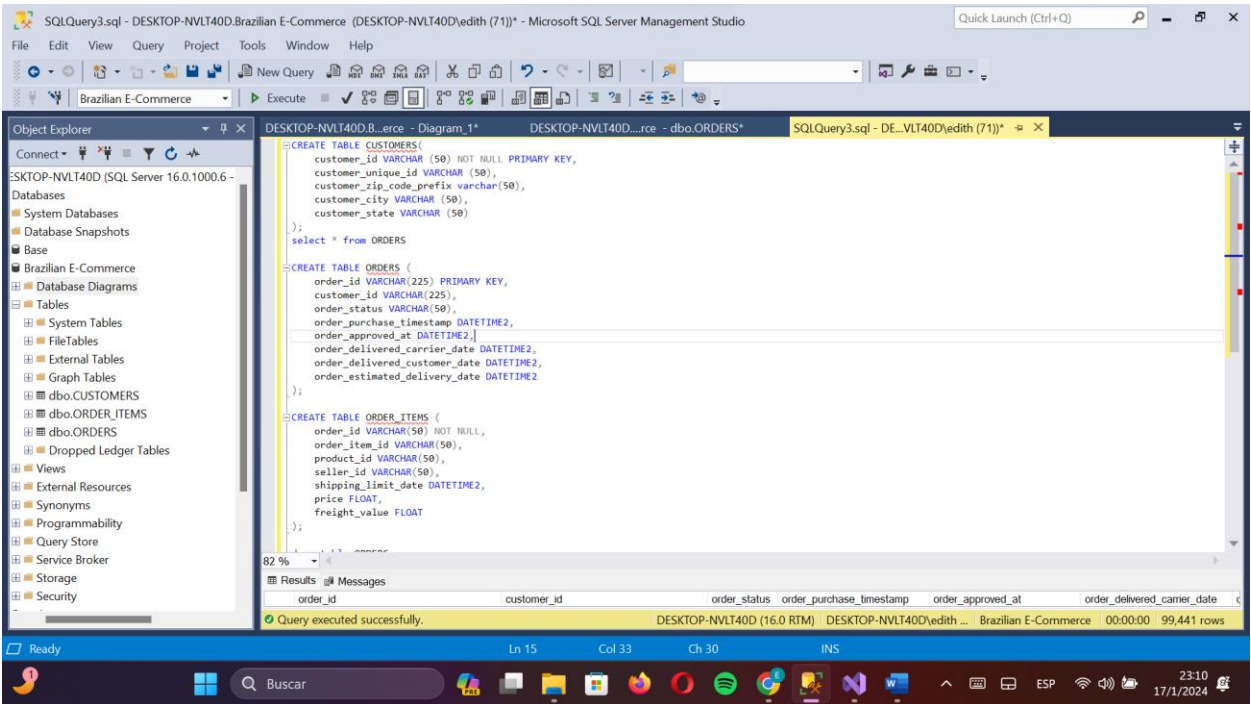


Figura 3. Creación de tablas SQL Server

Y relacionadas entre si con sus respectivas llaves primarias y foráneas:

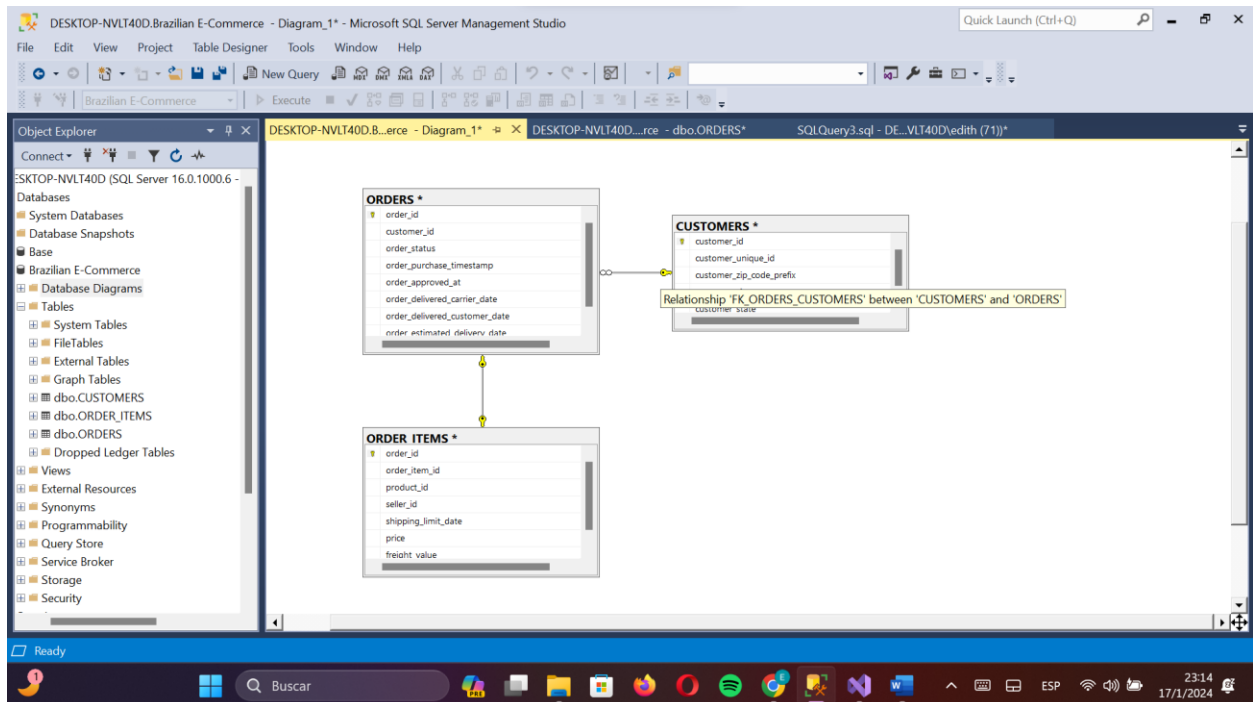


Figura 4. Diagrama de la base de datos

PARTE 2. AUTOMATIZACIÓN DEL PROCESO

- Para hacer la carga de los datos desde los archivos .csv hice uso del SQL SERVER IMPORT AND EXPORT WIZARD, inicié con la tabla CUSTOMER, aplicando las delimitaciones de caracteres según como se encontraban los datos en los archivos, aplicando la coma que separaba las columnas, y el salto que determinaba cada fila, y transformando el tipo de datos que venían desde el origen.

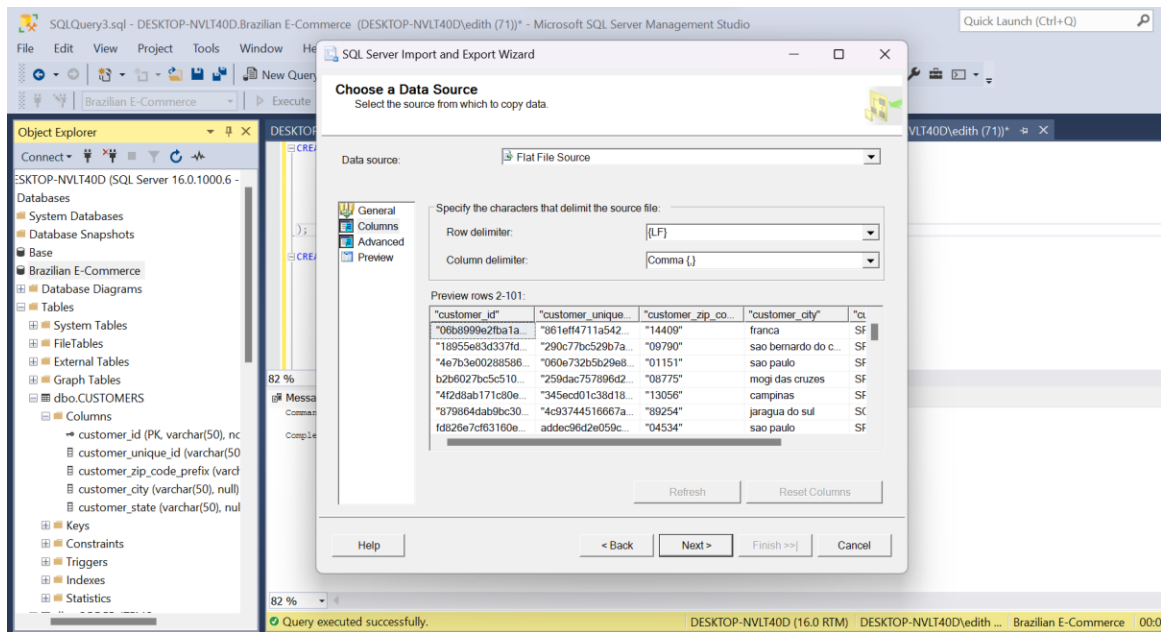


Figura 5. Importando datos de Excel a SQL Server

- Indicando la fuente de los datos que en este caso es el archivo customers_dataset, y el destino que sería la tabla CUSTOMERS previamente creada en la base de datos

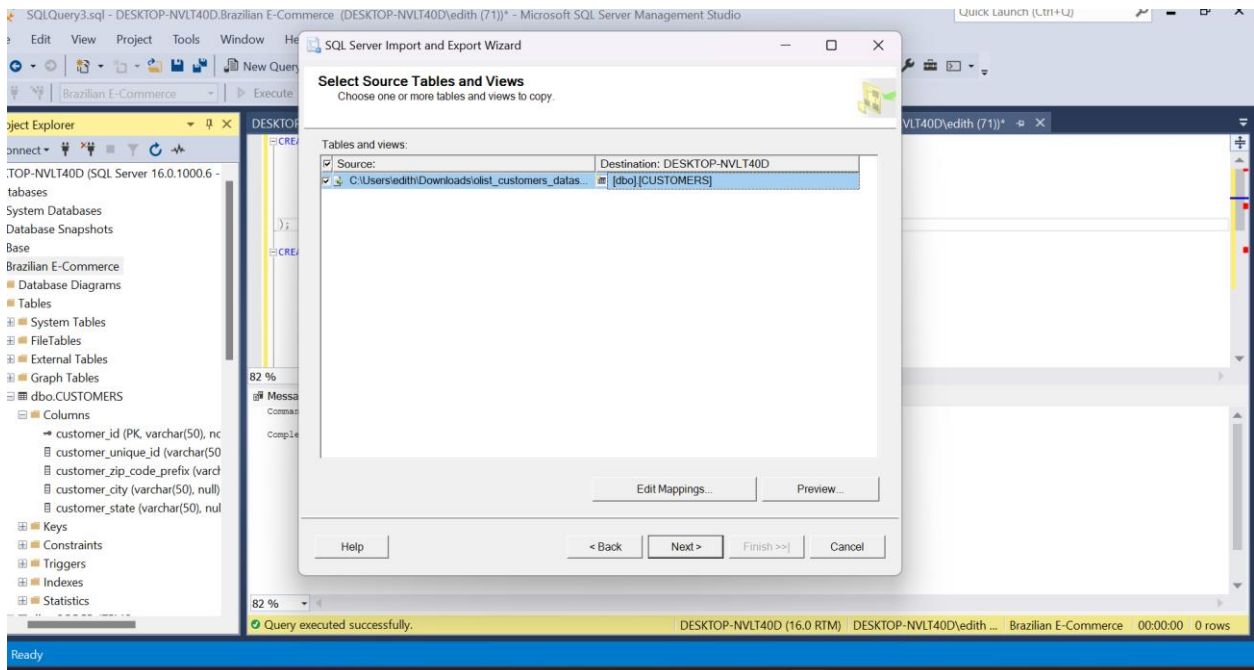


Figura 6. Importando datos de Excel a SQL Server

- Exportando los datos de manera exitosa:

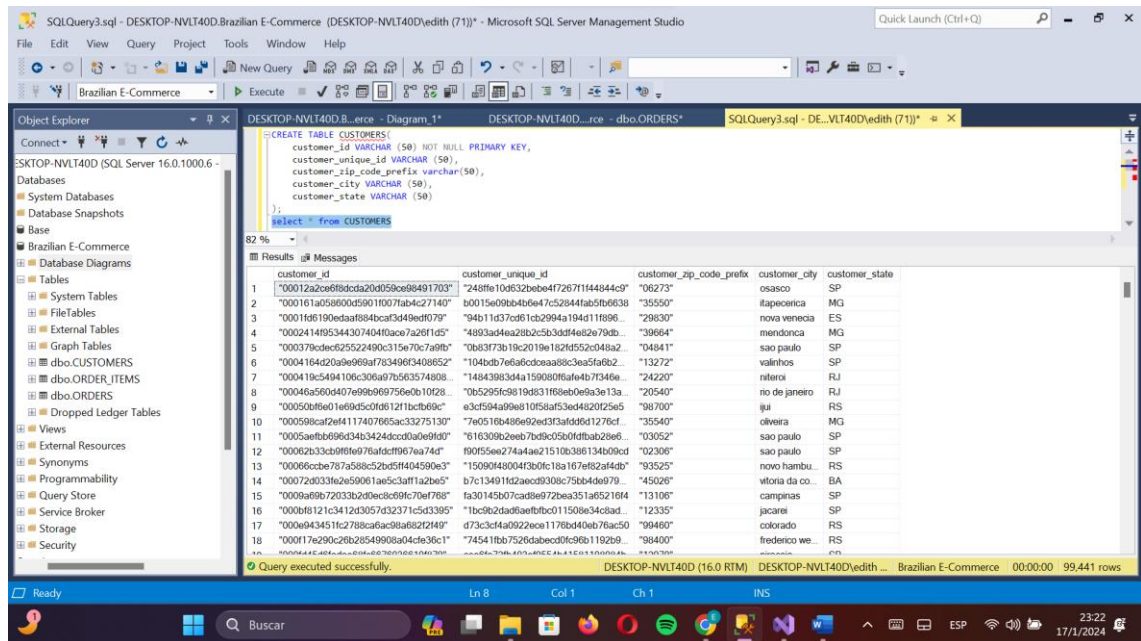


Figura 7. Select tabla Customers

- Al querer repetir estos pasos con las siguientes 2 tablas (Order_Items y Orders) el caso no fue el mismo, habían conflictos por el tipo de dato de las columnas destinadas a fechas, a pesar de cambiar el tipo de dato al querer mostrar los datos esta no los mostraba correctamente, por lo que decidí recurrir mejor al uso de Visual Studio y los proyectos Integration Services que permiten migrar datos de varios archivos .csv a una base de datos-ETL-SSIS por medio de los Data Flows, añadí uno para cada tabla

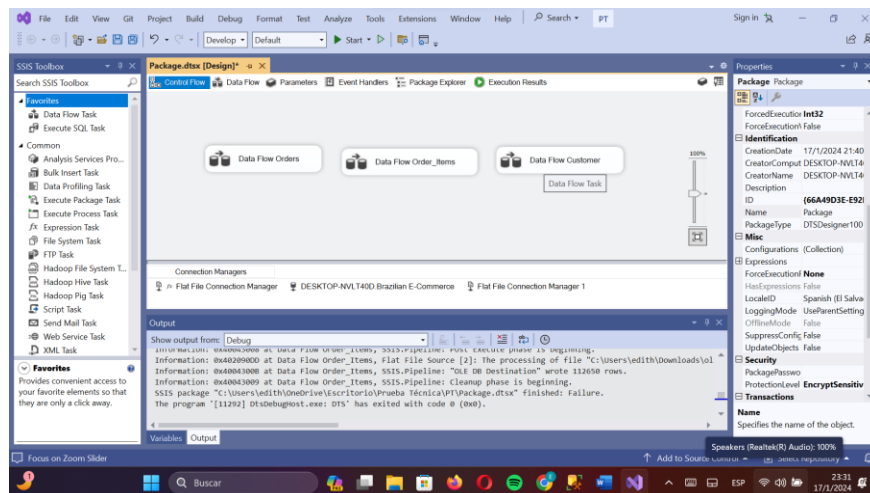


Figura 8. Creando Data Flows

- Para cada uno de estos seleccioné del área de Toolbox las herramientas para seleccionar el origen de los datos, en este caso un Flat File Source en el cual extraje los datos de los archivos .csv correspondiente a cada Data Flow, a continuación hice la transformación de los datos para ciertos campos no aceptar nulo como lo eran los campos de ID de cada tabla, transformación de los tipos de datos para que no hubiera conflicto a la hora de extraer los datos a SQL Server, mapeo de datos entre el origen y el destino etc. Realicé la conexión entre la fuente y el destino de los datos, los cuales al ejecutarlos no dieron error alguno.

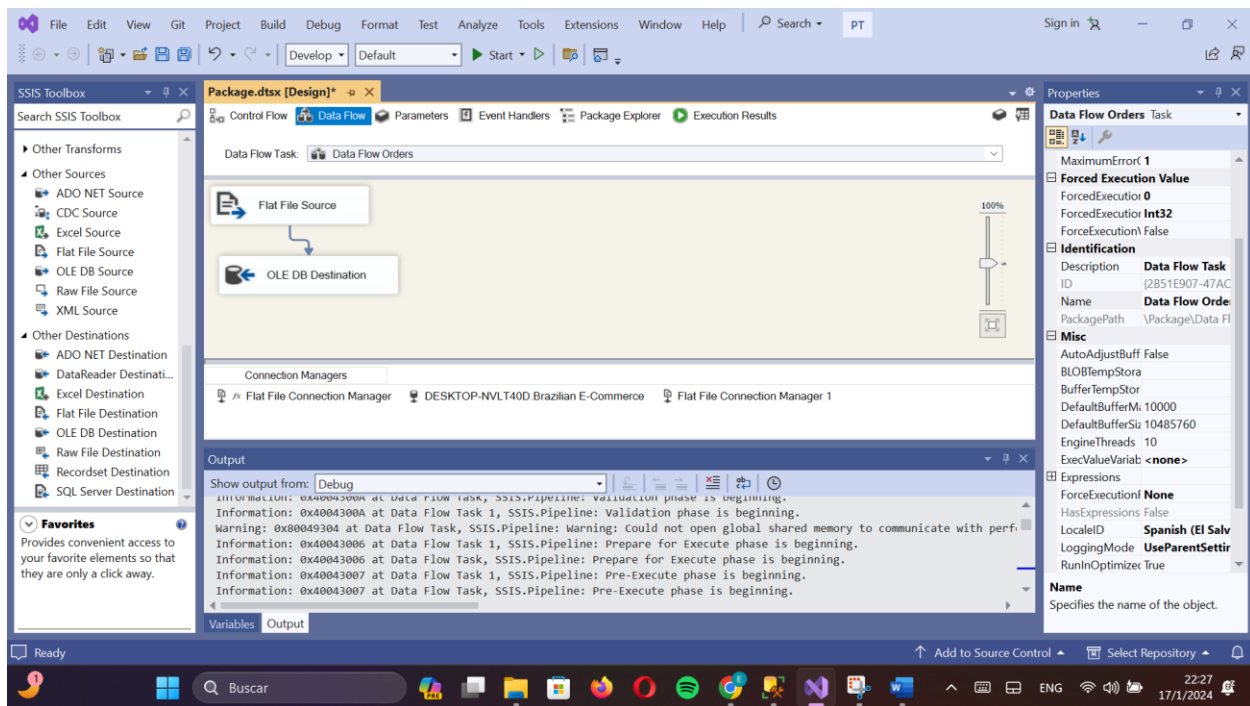


Figura 9. Conexión entre Fuente y Destino

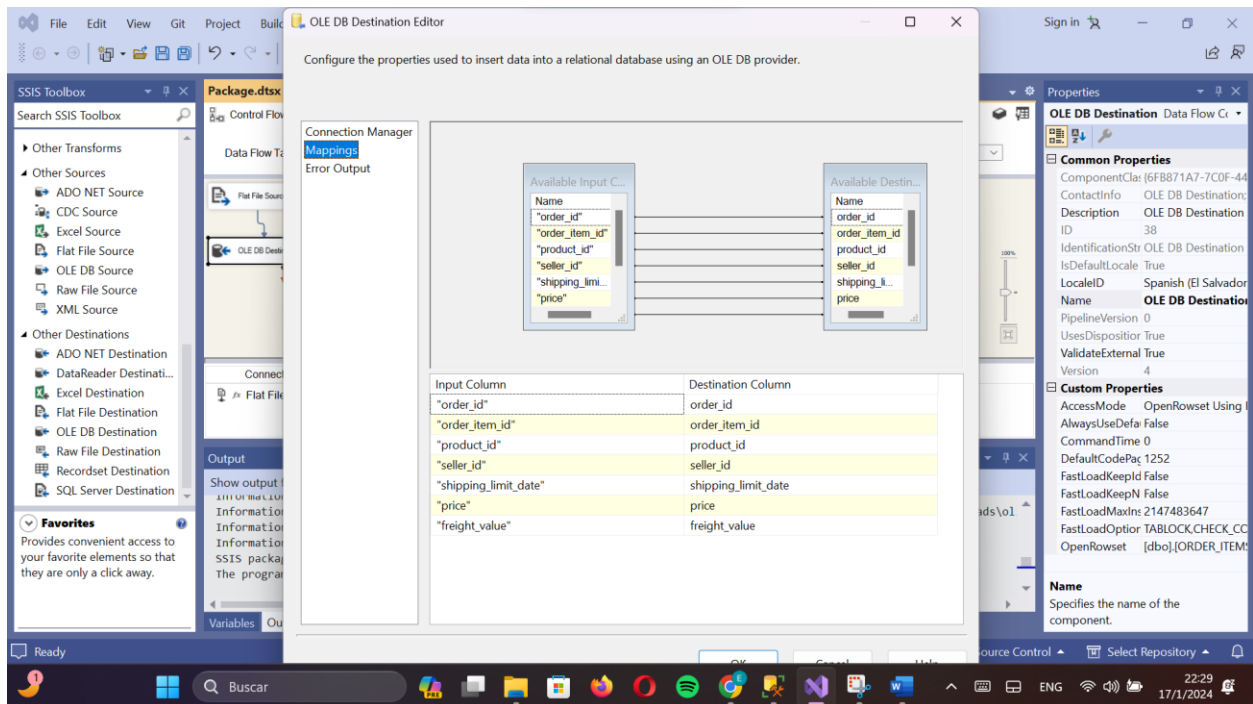


Figura 10. Mapeo de columnas de tabla ORDER_ITEM

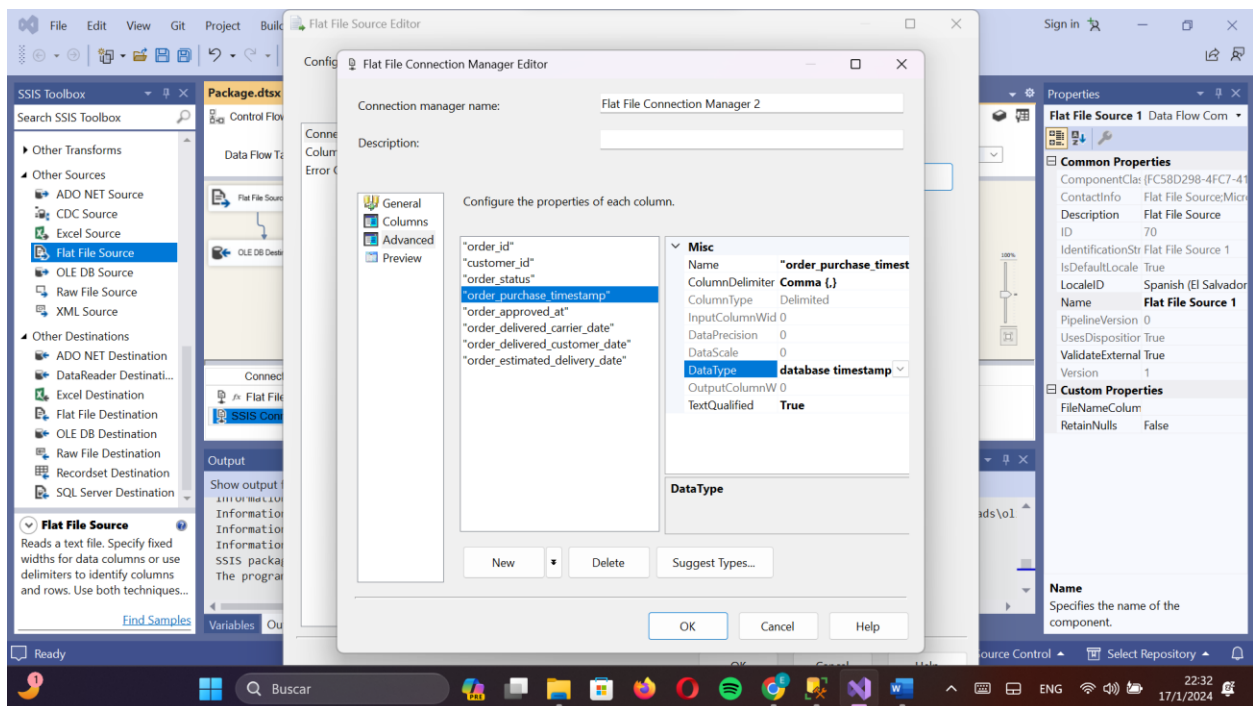


Figura 11. Transformación del tipo de datos

SQLQuery3.sql - DESKTOP-NVL40D.Brazilian E-Commerce (DESKTOP-NVL40D\edith (71)) - Microsoft SQL Server Management Studio

Object Explorer: Brazilian E-Commerce > Tables > dbo.ORDERS

```

CREATE TABLE CUSTOMERS (
    customer_id VARCHAR (50) NOT NULL PRIMARY KEY,
    customer_unique_id VARCHAR (50),
    customer_zip_code_prefix VARCHAR (50),
    customer_city VARCHAR (50),
    customer_state VARCHAR (50)
);
select * from ORDERS

```

order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date
1	"00010242f8c5a6d1ba2dd792cb16214"	delivered	2017-09-13 08:59:02.0000000	2017-09-13 09:45:35.0000000	2017-09-19 18:34.16.0000000
2	"00018f772f0320c557190d7a144bd33"	delivered	2017-04-26 10:53:06.0000000	2017-04-26 11:05.13.0000000	2017-05-04 14:35.00.0000000
3	"0002299c398224ef6ca065d44fc703e"	delivered	2018-01-14 14:33.31.0000000	2018-01-14 14:48.30.0000000	2018-01-16 12:36.48.0000000
4	"00024acbcd0a6daa1e931b038114c75"	delivered	2018-08-08 10:00.35.0000000	2018-08-08 10.10.18.0000000	2018-08-10 13:28.00.0000000
5	"00024b26cd5fd7ce69dfabb4e55b4fd9"	delivered	2017-02-04 13:57.51.0000000	2017-02-04 14.10.13.0000000	2017-02-16 09:46.09.0000000
6	"00048cc3ae777c65dbb7d2a0634bc1ea"	delivered	2017-05-15 21:42.34.0000000	2017-05-17 03:55.27.0000000	2017-05-17 11:05.55.0000000
7	"00054e6431b9d767580bcb819fb4a32"	delivered	2017-12-10 11:53.48.0000000	2017-12-10 12.10.31.0000000	2017-12-12 01:07.48.0000000
8	"0005786e39319847cbb9d288c5617fa8"	delivered	2018-07-04 12:08.27.0000000	2018-07-05 16.35.48.0000000	2018-07-05 12.15.00.0000000
9	"0005a1a1728cd9d785b8e2b08b904578e"	delivered	2018-03-19 18:40.33.0000000	2018-03-20 18.35.21.0000000	2018-03-28 00:37.42.0000000
10	"0005f50442cb953dcd1d21e1fb923495"	delivered	2018-07-02 13:59.39.0000000	2018-07-02 14.10.56.0000000	2018-07-03 14.25.00.0000000
11	"0006112a7b2c9da3e4d15a52dc8a4af1"	delivered	2018-03-24 22.16.10.0000000	2018-03-24 22.28.09.0000000	2018-03-25 02.56.55.0000000
12	"00063b381e2406b52ad429470734ebd5"	delivered	2018-07-27 17:21.17.0000000	2018-07-27 18.00.08.0000000	2018-07-30 14.52.00.0000000
13	"0006c9d6010a6e59a89b2c340bf65a7"	delivered	2018-07-24 17:04.17.0000000	2018-07-24 17.24.20.0000000	2018-07-25 11:02.00.0000000
14	"000828aa423d2a3f00fb17cd7d8719"	delivered	2018-02-13 22.10.21.0000000	2018-02-15 03:55.52.0000000	2018-02-20 18.23.28.0000000
15	"0009792311464db532f7f85bf7b182ae"	delivered	2018-08-14 20:43.09.0000000	2018-08-15 12.15.10.0000000	2018-08-15 15.28.00.0000000
16	"0009c3a17916a706d71784483a5d643"	delivered	2018-04-25 09.10.41.0000000	2018-04-25 09.33.19.0000000	2018-04-27 15.14.00.0000000
17	"000a2e2e5dbad2f9dd70584c5ad2ed"	delivered	2018-05-11 20:33.38.0000000	2018-05-11 20:57.03.0000000	2018-05-16 14.26.00.0000000

Query executed successfully. DESKTOP-NVL40D (16.0 RTM) DESKTOP-NVL40D\edith ... Brazilian E-Commerce 00:00:00 99,441 rows

Figura 12. Datos Cargados correctamente tabla ORDERS

SQLQuery3.sql - DESKTOP-NVL40D.Brazilian E-Commerce (DESKTOP-NVL40D\edith (71)) - Microsoft SQL Server Management Studio

Object Explorer: Brazilian E-Commerce > Tables > dbo.ORDER_ITEMS

```

CREATE TABLE CUSTOMERS (
    customer_id VARCHAR (50) NOT NULL PRIMARY KEY,
    customer_unique_id VARCHAR (50),
    customer_zip_code_prefix VARCHAR (50),
    customer_city VARCHAR (50),
    customer_state VARCHAR (50)
);
select * from ORDER_ITEMS

```

order_id	order_item_id	product_id	seller_id	shipping_limit_date	price	freight_vat
1	"050f93440d6b604a9d9c01b12d9d455"	"112e185664c6762c1ca6058d072633"	"440d6ab244315c632130ecfb63827b1"	2017-10-17 04:56.16.0000000	109.9	13.11
2	"05104976125ea1e5cc04bc1f506a436a"	"63bfa27a0ce021fd896dc2758b6d656"	"612170e34b07004b3ba37ae81836b4c"	2018-08-19 22:44.03.0000000	198.9	18.32
3	"05108247830e6ab6f04da38aff20907f"	"a645f291d812aa4d85f034a5c469dec"	"343e2916d76a3748b0698980efbaa294e"	2018-04-25 12:35.24.0000000	22.9	7.39
4	"0510b3fb25b06b2b1221ec1cd2e495ee"	"e306ca54c91b211392317d5b4632c9f6"	"e49c26c3ed4add6227d5121a6b6e4d37"	2017-06-15 20:35.13.0000000	135.3	44.8
5	"0510632b86f01064ebbd6f7d6b63bb5"	"10070e299297c7d609ee32011236b8e"	"8a130737016838139d31876787a39c9"	2018-06-29 00:31.25.0000000	95.2	18.47
6	"0510b6f17088fbb22eed8b459322c62"	"e446f05803a3a2453ec36421e06f0f"	"730937b188cb151eb2eb4a96a42c113"	2018-04-30 19:55.14.0000000	105.3	7.92
7	"05123197f45cd213617dbd9523bd23f7"	"110592734fca1a29377f7e9bc89807ce2"	"431af27296bcb0519d890aa5a05b621"	2017-07-20 11:43.05.0000000	120.9	15.6
8	"0512553c96f9d59a9572ce7eebac43a"	"1acb187689c5489d5f76abc9881165e"	"6d803bc79cc31e41c4c789a575933bc7"	2017-11-06 03:06.45.0000000	135	13.97
9	"0512bcd90b610913b8177d5324c5ee0d"	"68bf2e767db9a46042e54da87399ca"	"1acd49d1dc41f1589fc2b934da68bf2"	2017-06-30 16:45.18.0000000	69.9	17.74
10	"05147da98ae779803eaa2a5490010a9"	"0a29c0e1f89190cd32d9f6a23f7e9a8c"	"138dbae45cd2f1e24378131a6801526"	2018-06-14 01:55.12.0000000	9.9	7.39
11	"0514e4dbba025db4265c730bf7371261"	"e343a7b068e928c3f9f0e09c234abc0"	"f0837c8d71434931d9c38e7679234797"	2018-04-24 19:50.48.0000000	37	22.06
12	"0516c22655fe1f0b3390f4ee3c1c6f"	"67c3b207e496408eb0f63ae0c234abc0"	"7c67e1448b00f6e96d365c0a6b010ab"	2017-02-09 11:49.23.0000000	29.99	15.56
13	"0517a3e68dac33089959e2c2144db36e"	"53b3e6df7eb7c4158e9d5468772e08"	"7d13fca15225358621be4086a1eb0964"	2018-05-03 09:11.40.0000000	99.9	0
14	"0517b06346d2d436525fd79ef72c0e"	"d1c42706a0f73f6b889a5c70b1f2ac4"	"a1043baf471dff536dc0462352be48"	2018-03-21 00:10.59.0000000	129	58.76
15	"0517b06346d2d436525fd79ef72c0e"	"d1c42706a0f73f6b889a5c70b1f2ac4"	"a1043baf471dff536dc0462352be48"	2018-03-21 00:10.59.0000000	129	58.76
16	"0517edc6d6ee491e072a844fb8b54c86"	"93f1c0e4b913dbf3213569661cb978bc"	"b9ca8ebaa564aa036394a700f63e69f"	2018-05-04 03:11.13.0000000	38.99	8.88
17	"0517f595d01e0c246ac759a754dc6a64"	"2b460f9f84b8e18874494203496bc310"	"cc419e065a3c5ba77189a1882b7556a"	2017-10-23 17:08.39.0000000	89.99	12.97

Query executed successfully. DESKTOP-NVL40D (16.0 RTM) DESKTOP-NVL40D\edith ... Brazilian E-Commerce 00:00:02 337,950 rows

Figura 12. Datos Cargados correctamente tabla ORDER_ITEMS

BIBLIOGRAFÍA

1. <https://learn.microsoft.com/es-es/sql/integration-services/data-flow/integration-services-data-types?view=sql-server-ver16>
2. <https://www.sqlshack.com/importing-and-working-with-csv-files-in-sql-server/>