

Experiencia de aprendizaje

Desafío 1

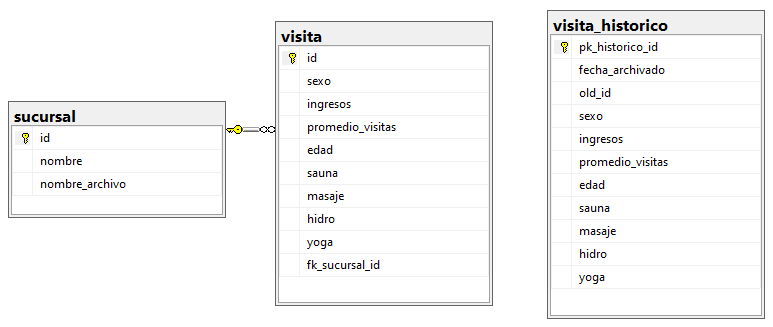
Desarrollado por **Edgar Mauricio Rivas Hernández** | RH131925

Herramientas utilizadas:

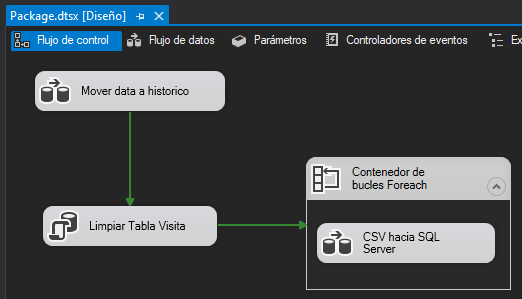
* Visual Studio 2019
* SQL Server 2019 Developer Edition
* SQL Server Management Studio 18
* SSIS v3.12

# Ejercicio 1: Spa Diego

Se diseñó la siguiente base de datos para almacenar la información proporcionada por el cliente. Se separó sucursal en su propia tabla para poder utilizarla como filtro desde el ETL. Se crea una tabla visita\_historico para almacenar la información actual en de la tabla visita antes de cargar nueva data.

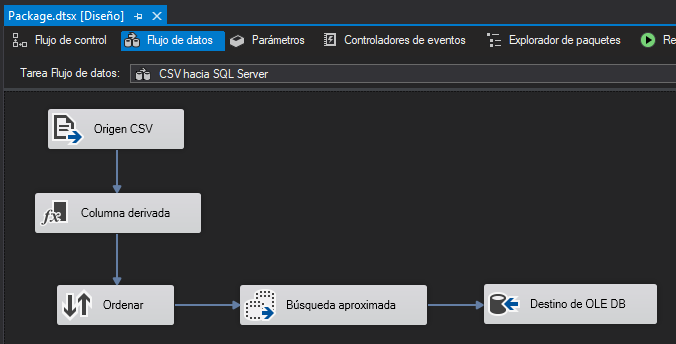


Vista general del Flujo de control: Se agrega al inicio una Tarea de flujo de datos “Mover data a histórico”, que se encarga de mover la data de visitas a visitas\_historico. “Limpiar Tabla Visita” ejecuta un truncate sobre la tabla visitas. El Contenedor Foreach procesa cada archivo CSV y lo envía a la tarea de flujo principal “CSV hacia SQL Server”



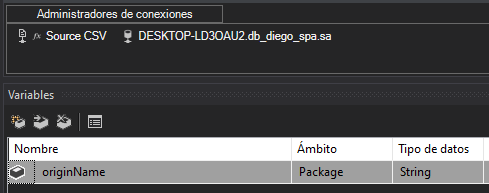
Vista del flujo “CSV hacia SQL Server”. Contiene los siguientes componentes:

1. Origen CSV, utiliza una expresión para tomar el path del archivo del ForEach mostrado anteriormente
2. Columna Derivada, agrega el nombre de archivo como columna a todos los registros
3. Ordenar, ordena los registros por sucursal
4. Búsqueda aproximada, consulta la tabla sucursales, su campo nombre\_archivo para verificar a qué sucursal se debe asignar. Agrega la columna fk\_sucursal\_id
5. Destino de OLE DB, escribe a la base db\_diego\_spa

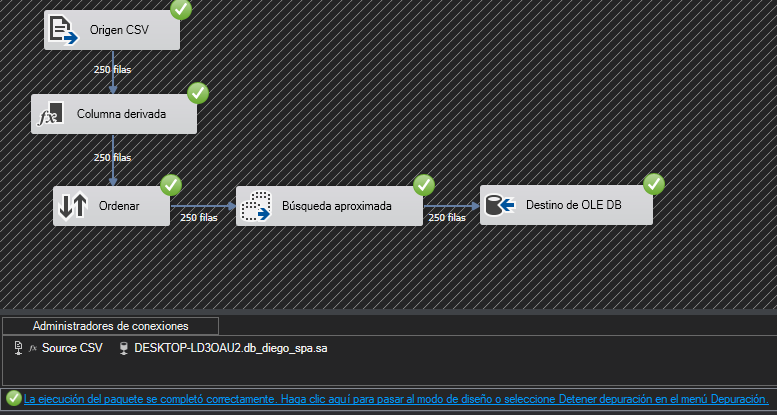


Vista de las conexiones y variables.

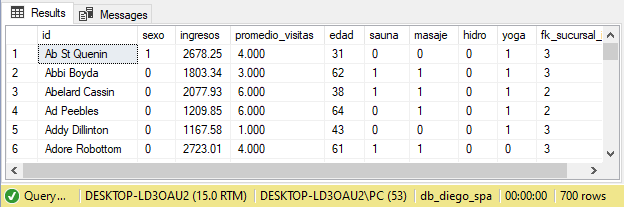
1. Source CSV, la fuente de archivos planos usa una expresión basada en la variable originName
2. DESKTOP-LD3OAU2.db\_diego\_spa.sa, la conexión al servidor SQL Server
3. originName, es la variable donde el ForEach deposita el path del archivo que se lee actualmente



Muestra del paquete en ejecución

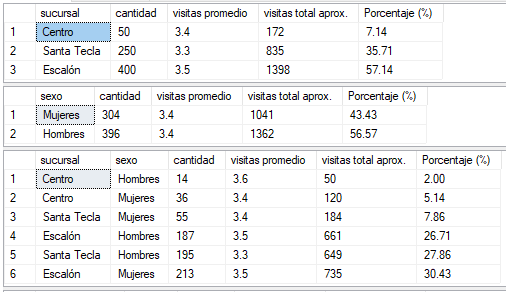


Consultando la tabla vistas, vemos los 700 registros esperados

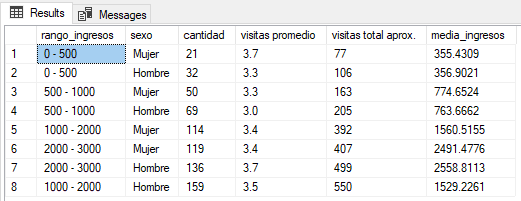


Consultas para agrupar clientela.

1. Clientes / Visitas totales por sucursal
2. Clientes / Visitas totales por sexo
3. Mezcla de las anteriores



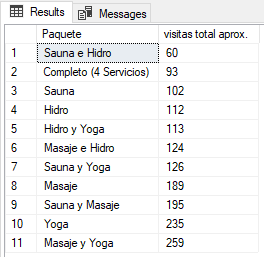
1. Por rango de ingresos, mostrando ingresos promedio.



Se puede detectar con esto los siguientes puntos a considerar:

1. Hay más clientela masculina
2. Los clientes que ganan más de $1000 conforman casi dos terceras partes de la clientela
3. La sucursal más visitada es Escalón, en contraste la sucursal Centro casi no tiene visitas
4. Las mujeres que ganan menos de $500 conforman el grupo más pequeño de la clientela (21), y los hombres que ganan entre $1000 y $2000 son el grupo más grande (159)

Otra consulta de interés es el uso de servicios, podemos ver que el servicio más usado por si solo es Yoga. Y el que menos clientes usan es el Sauna.



# Ejercicio 2

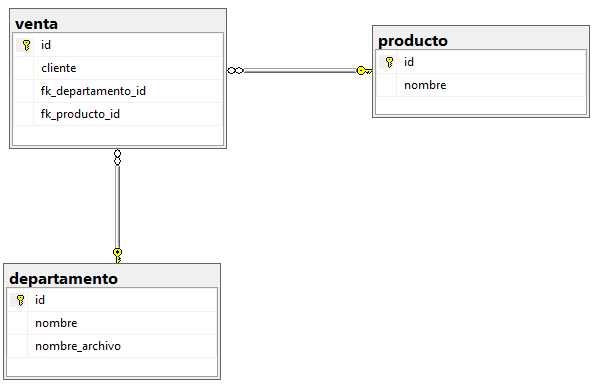
Nota: este ejercicio contiene dos soluciones, inicialmente se planteó crear una tabla con todas sus dimensiones incluidas, excepto por el departamento, pero esta versión resultó inviable para realizar consultas. Opté por realizar un segundo intento agregando la dimensión “producto” para facilitar el análisis de los resultados.

Todos los archivos del primer intento se encuentran dentro de la carpeta Ejercicio2

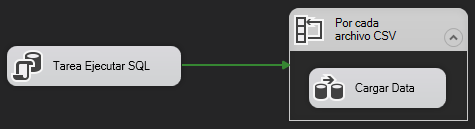
## Desarrollo de segundo intento



Para este ejercicio desarrollé la siguiente base de datos, donde la tabla de hecho es “venta”. Las tablas “producto” y “departamento” corresponden a dimensiones de la venta.

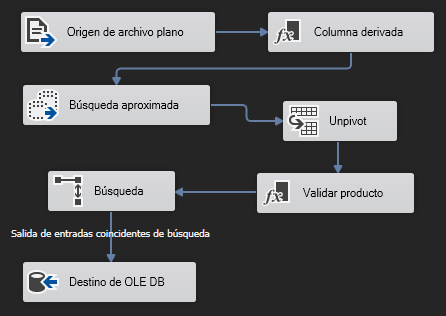


El flujo de control es el siguiente, iniciando por un truncado de la tabla venta, para ingresar la nueva data. Un contenedor ForEach para cada archivo CSV y el Flujo de datos “Cargar data”

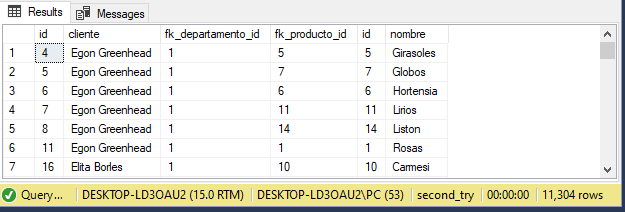


El flujo de datos contiene estos componentes:

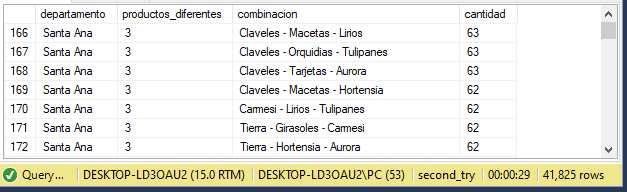
1. Origen de archivo plano para los archivos CSV
2. Columna derivada para agregar el path de archivo como columna
3. Búsqueda aproximada para hacer match entre el nombre de archivo en base de datos y el path almacenado en la columna en el paso anterior
4. Anulación de dinamización “Unpivot”, sirve para convertir todas las columnas de producto en registros de la tabla venta, a partir de aquí se hace un loop a través de cada columna afectada
5. Columna derivada para agregar columna “producto” en la cual se guarda el nombre de la columna a procesar en ese momento, ejemplo producto=Rosas
6. Búsqueda, para hacer match exacto entre el nombre de la columna a procesar y el producto guardado en la tabla “producto”
7. Guardar un registro en ventas por cada columna de producto a procesar.



El resultado en la base de datos es el siguiente (consulta: select \* from [dbo].[venta] v inner join [dbo].[producto] p on p.id = v.fk\_producto\_id)

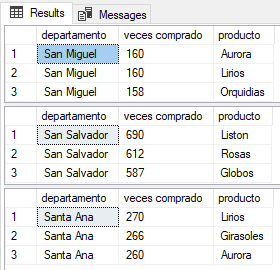


Utilizando una consulta recursiva se obtiene todas las combinaciones de productos y la cantidad de veces que se repite dicha combinación, como se puede observar, la cantidad de combinaciones hace inservible el resultado. Hablamos de poco menos de 42 mil combinaciones.



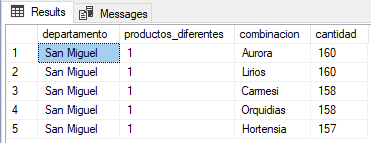
Para limitar este resultado y volverlo útil se utiliza algunas consultas para determinar algunos datos de interés

1. Top 3 productos en por departamento



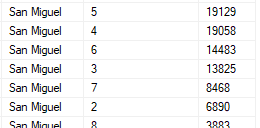
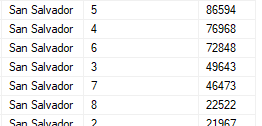
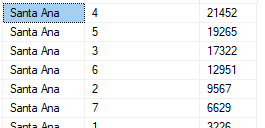
1. Top 5 combinaciones por departamento, se observa una tendencia a comprar por separado







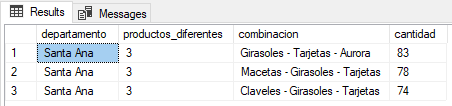
1. Consultar cantidades de repeticiones por número de productos diferentes para saber cuántos productos compra un cliente en promedio, muestra de San Salvador como referencia. Encontramos un dato de mucho interés, en los 3 departamentos, aunque no en el mismo orden, las ventas incluyen más veces 3, 4, 5 y 6 productos. Nos podemos enfocar en ese detalle para el último análisis



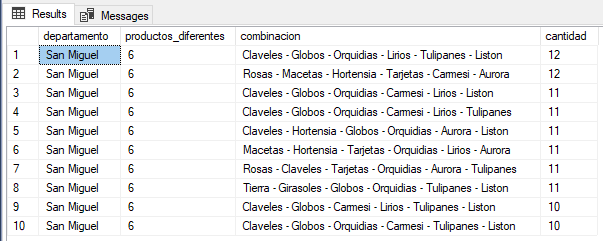
## Análisis final Ejercicio 2

Para entregar un informe a la floristería Fiorella se puede consultar por departamento, las combinaciones más usadas para 3, 4, 5 y 6 productos. Ejemplos

1. Top 3 combinaciones de 3 productos para Santa Ana



1. Top 10 combinaciones de 6 productos para San Miguel



## Conclusión

Con la base de datos y el proceso de integración de datos que se ha creado se puede realizar consultas muy sencillas con las cuales se determine la cantidad de productos que se consume por departamento y también por sus posibles combinaciones.