

UNIVERSIDAD DON BOSCO
INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN
DATAWAREHOUSE Y MINERÍA DE DATOS
CICLO 02-2020



Entrega del primer desafío practico

PRESENTADO POR:

GIRÓN CORNEJO, EMMANUEL ANTONIO GC180313
EDGARDO ANIBAL ZEPEDA LÓPEZ ZL180073

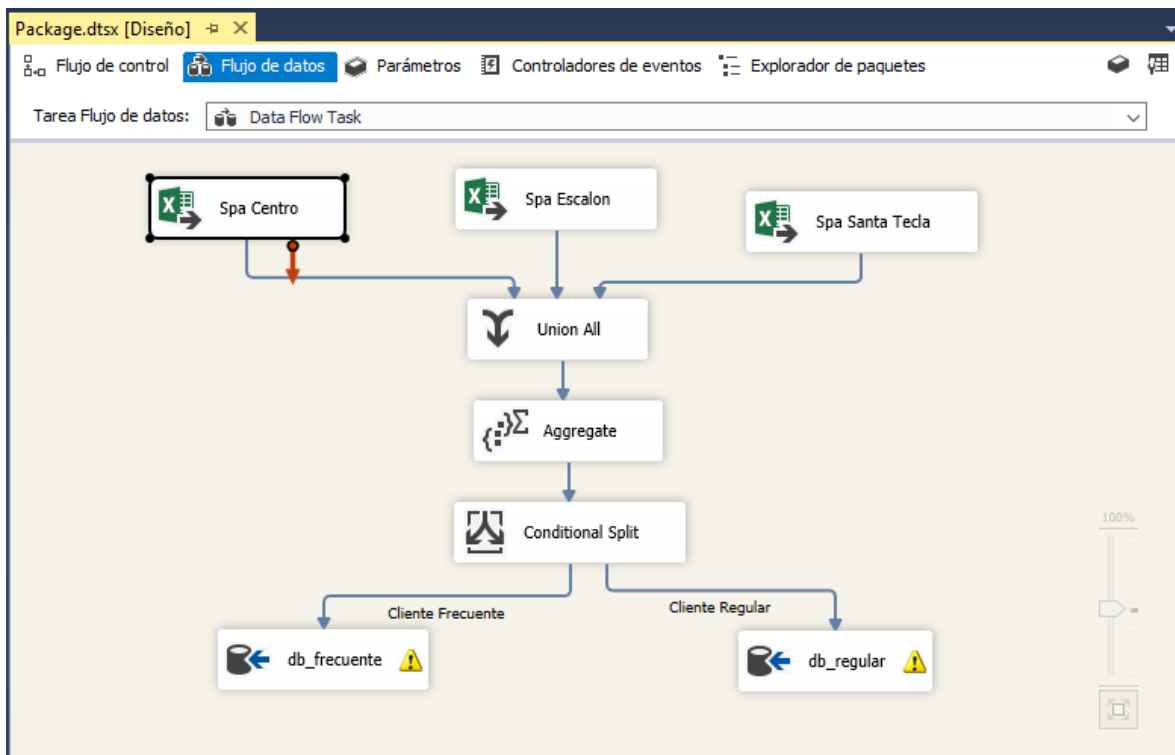
DOCENTE: Alexander Sigüenza

Grupo: 01L

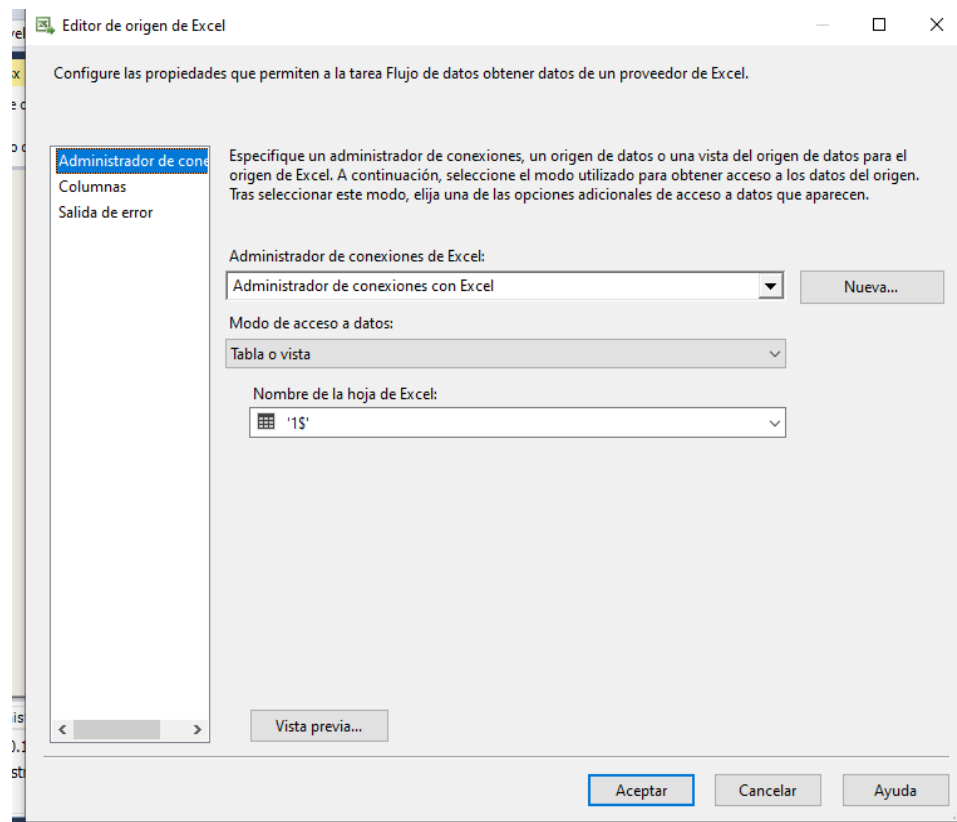
Soyapango, 28 de agosto de 2020

Ejercicio 1

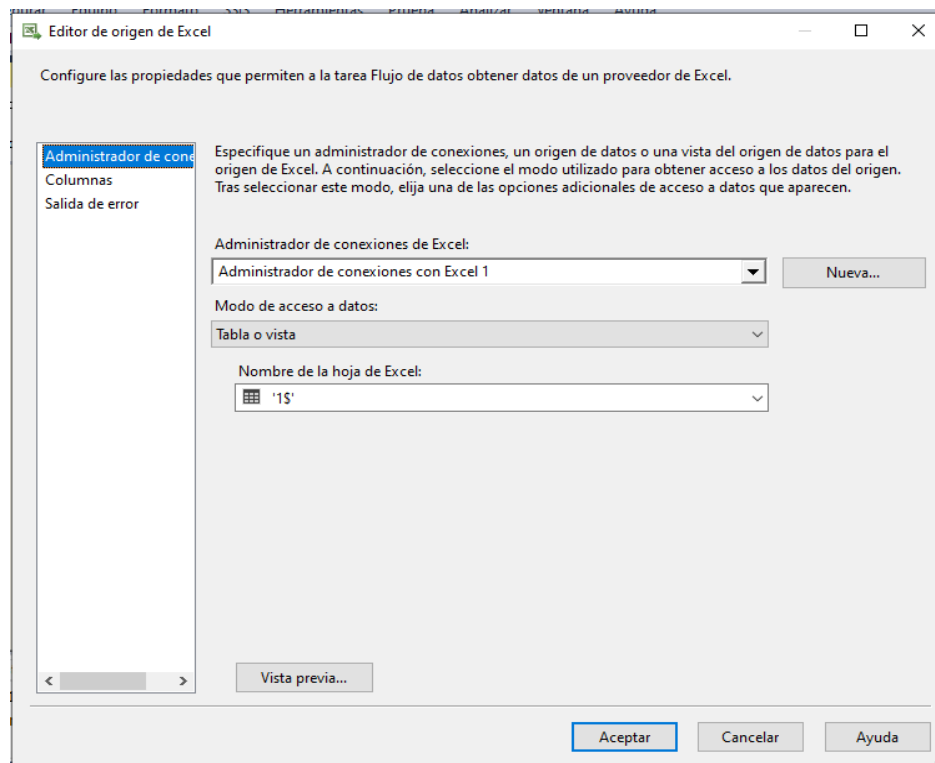
Primero configuramos la estructura general de nuestro ETL



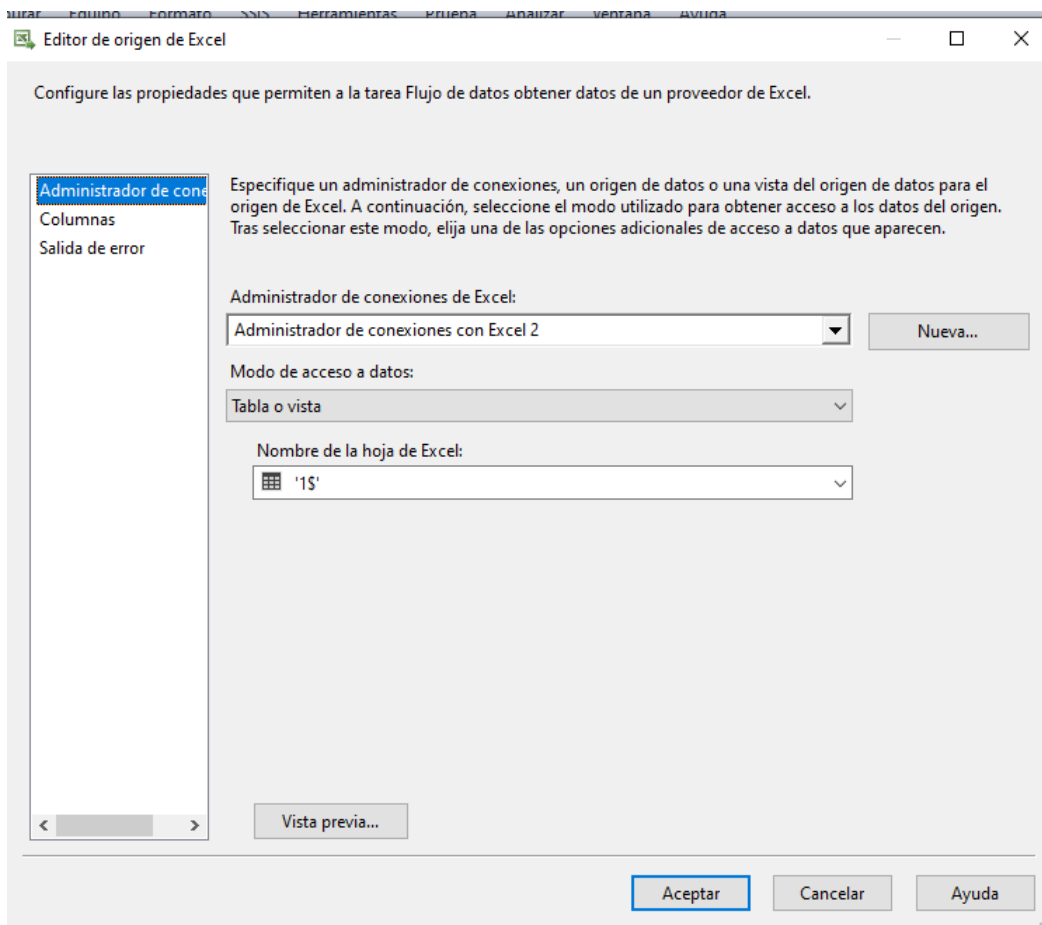
Se prosigue en la configuración del primer origen de datos:



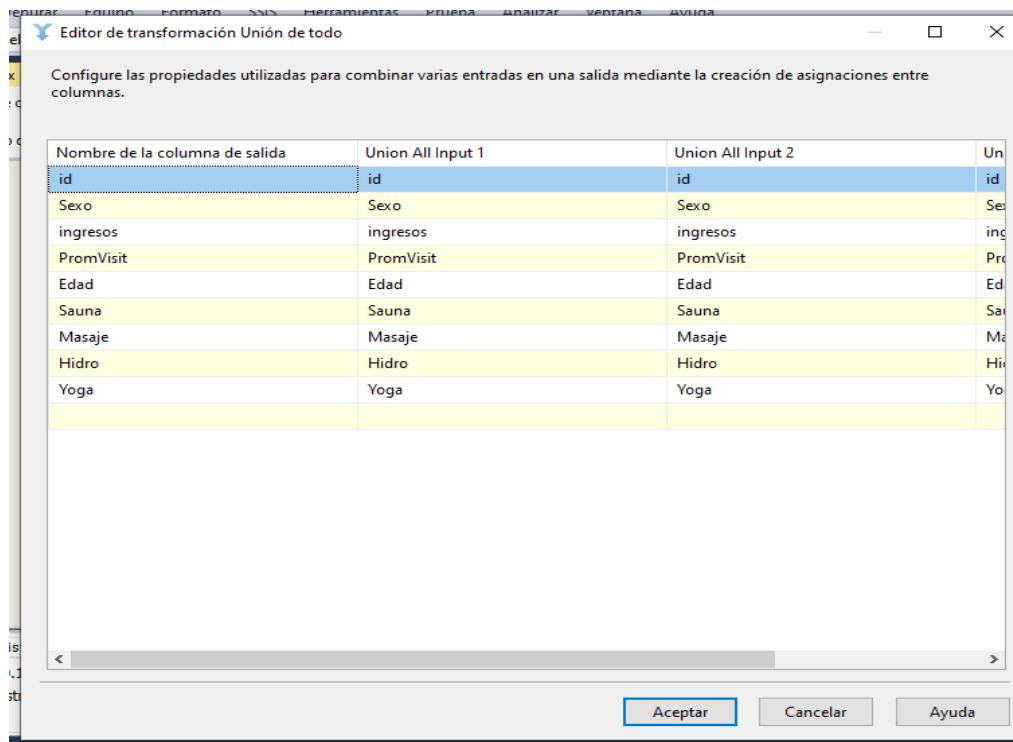
Se continua con el segundo origen de datos



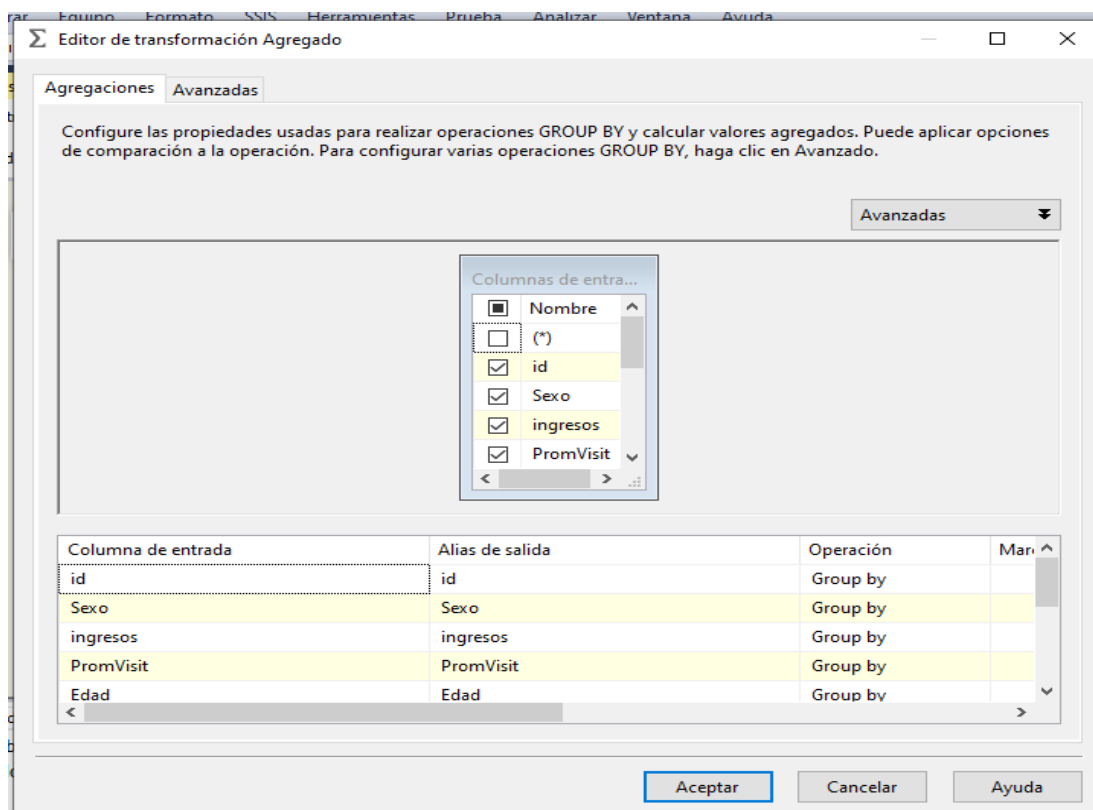
Y por ultimo el tercer origen



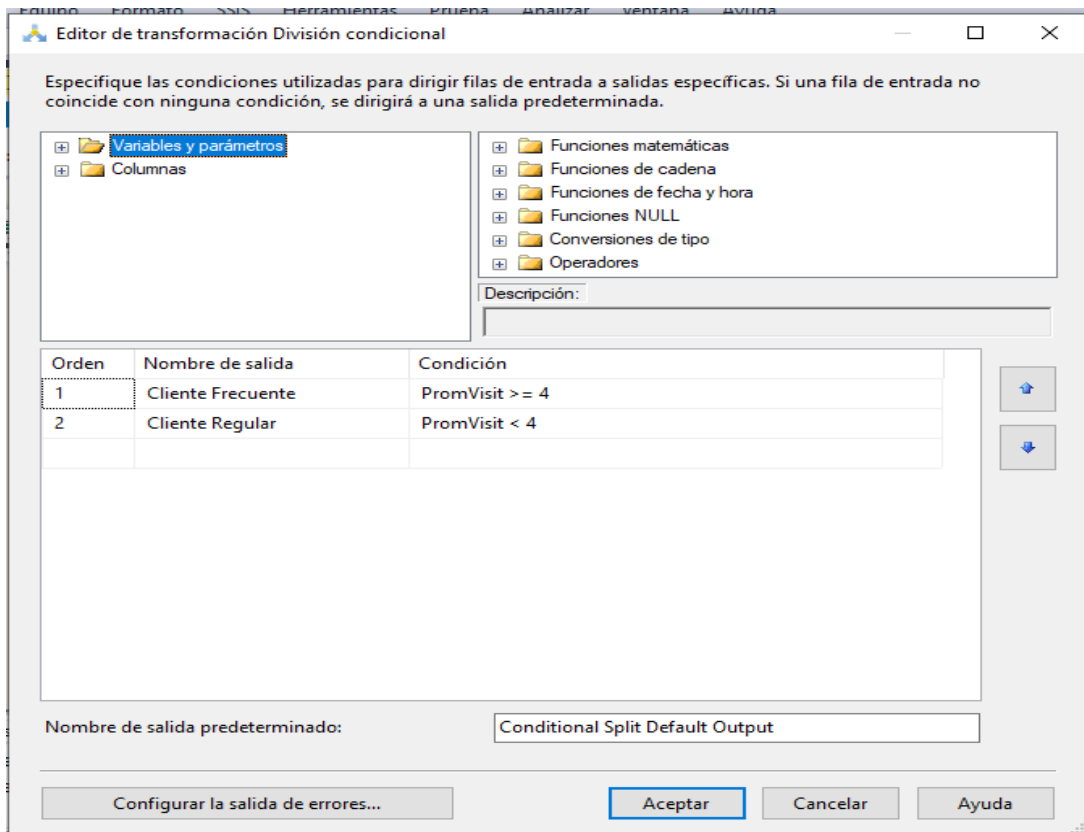
Continuamos con la configuración del Union All



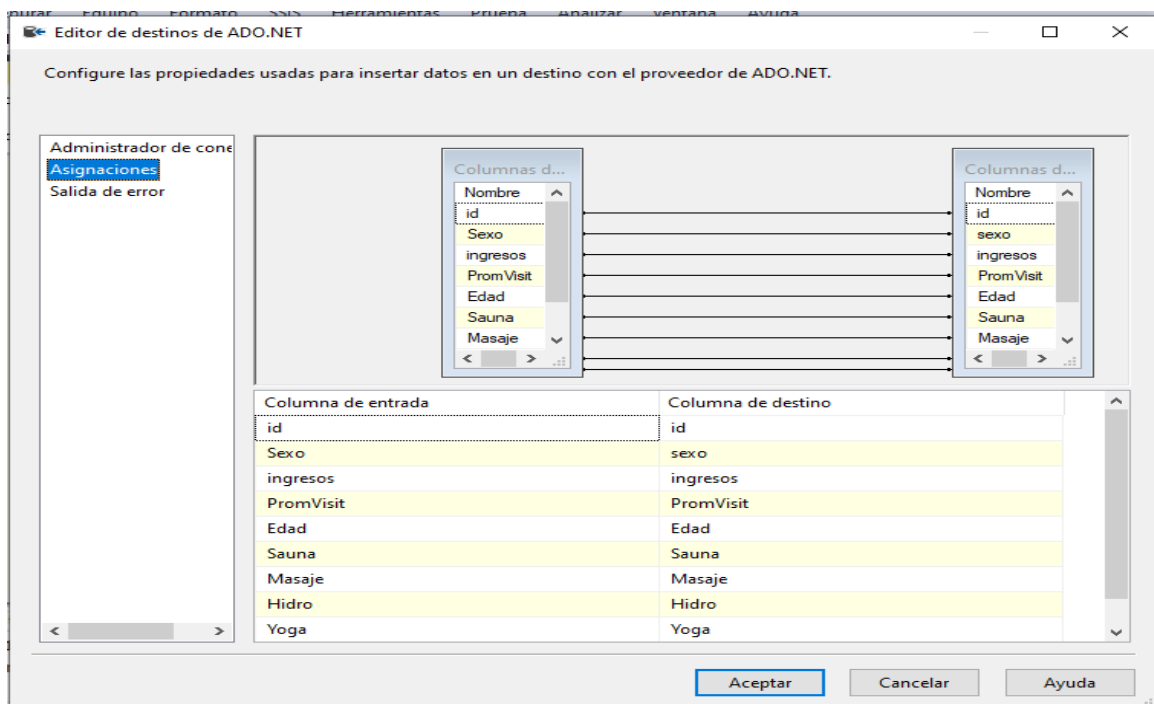
En esta parte agrupamos los datos



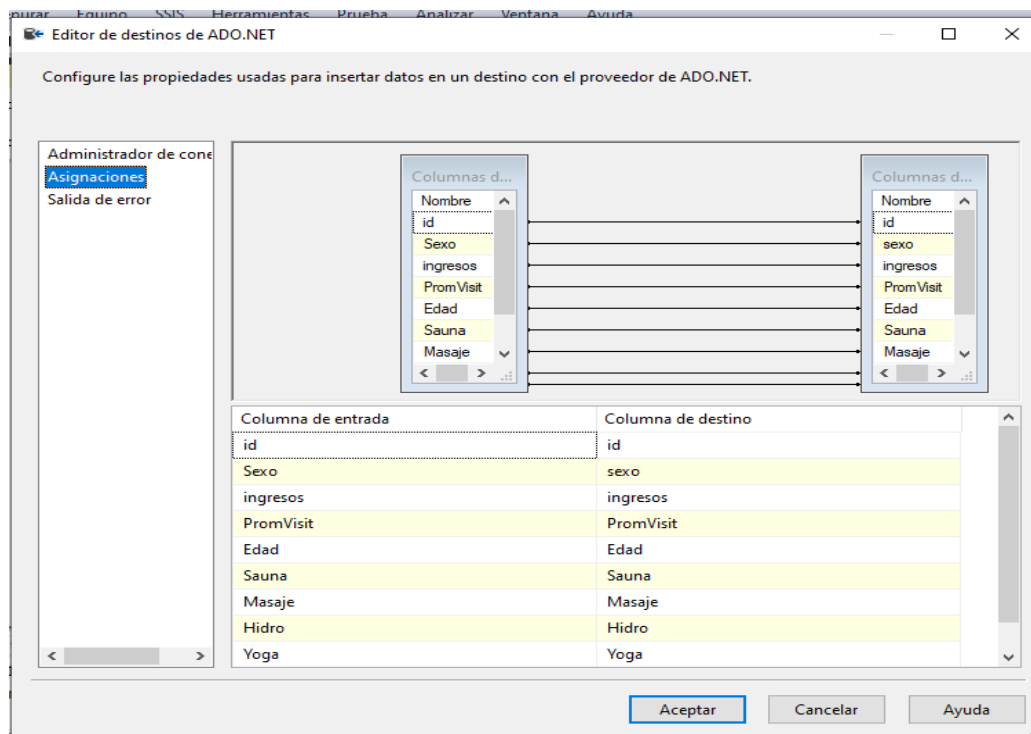
Para la configuración del conditional Split, hemos decidido que dividiremos los datos en 2 grupos, para el promedio de visitas mayor a 4 será un cliente frecuente, y para el menor un regular:



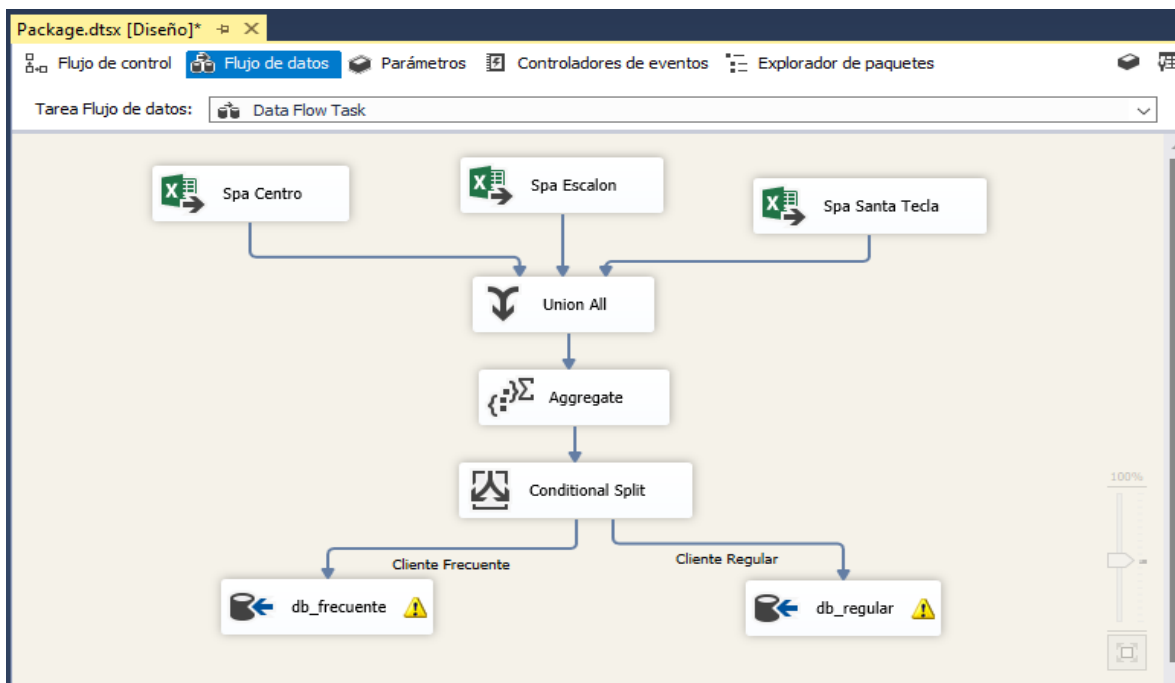
Configuramos en primer destino de datos correspondiente a la base de datos de los clientes frecuentes, en su respectiva tabla: frecuente

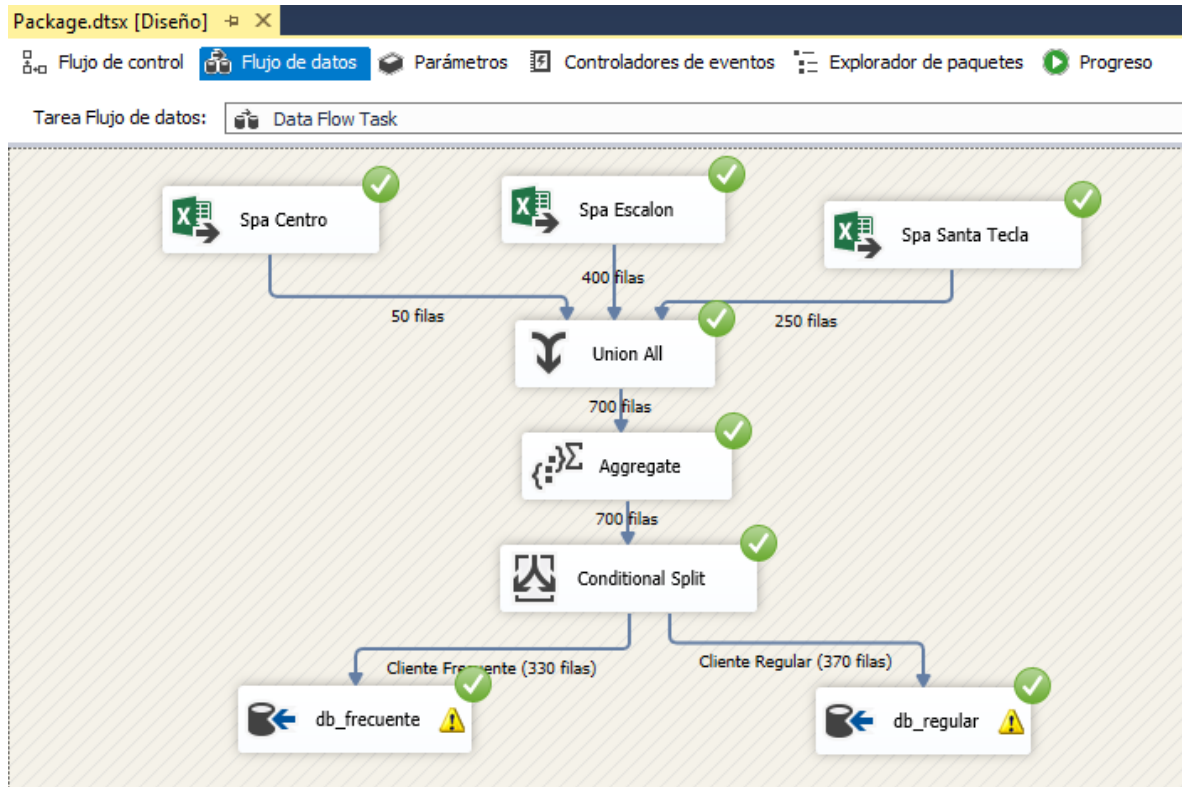


Se prosigue a configurar el segundo destino de datos, correspondiente a la base de los clientes regulares y en su respectiva tabla: regular



Proseguimos con la ejecución:





Ahora vemos el resultado en las bases de datos:

db_frecuente y regu...G6.master (sa (55)) X db_floristeria.sql - L..._floristeria (sa (53))*

```

create database db_clientes_frecuentes
use db_clientes_frecuentes

create table frecuente
(
  id varchar(50),
  sexo int,
  ingresos float,
  PromVisit float,
  Edad int,
  Sauna int,
  Masaje int,
  Hidro int,
  Yoga int
)

create database db_clientes_regulares
use db_clientes_regulares

create table regular
(
  id varchar(50),
  sexo int,
  ingresos float,
  PromVisit float,
  Edad int,
  Sauna int,
  Masaje int,
  Hidro int,
  Yoga int
)
  
```

```
select * from regular
```

```
use db clientes frecuentes
```

100 %

Results Messages

	id	sexo	ingresos	Prom Visit	Edad	Sauna	Masaje	Hidro	Yoga
1	Jenilee Aronov	0	466.33	3.74	58	1	1	0	1
2	Dorolisa Ronchetti	0	1499.1	1.94	39	0	0	0	0
3	Bernadette Jeannequin	0	982	1.95	36	0	1	1	0
4	Antonin Rolfo	1	1862.99	2.9	28	0	0	0	1
5	Dusty Mariette	0	1784.05	3.24	47	0	1	1	0
6	Tracee Blaydon	1	673.92	3.51	51	1	0	0	0
7	Wanda Micheli	1	977.44	1.37	50	0	1	0	0
8	Debora MattiCCI	0	565.25	1.99	35	0	0	0	0
9	Nevins Devany	0	1454.02	3.53	38	0	0	0	1
10	Geraldine Shoosmith	1	2431.95	3.03	28	1	0	1	0
11	Maddi Bufton	0	1658.62	1.26	30	1	0	0	1

Query executed successfully. LAPTOP-P1HT7PG6 (11.0 RTM) sa (55) db_clientes_regulares 00:00:00 370 rows

```
use db_clientes_frecuentes
```

```
select * from frecuente
```

100 %

Results Messages

	id	sexo	ingresos	Prom Visit	Edad	Sauna	Masaje	Hidro	Yoga
1	Tris Keough	1	1092.25	5.67	25	1	0	0	0
2	Christin Gillcrist	1	2274.88	5.24	40	1	1	1	0
3	Carlyn Texton	0	2027.09	5.5	59	0	0	0	0
4	Cointon Tigner	1	1709.8	6.83	64	1	1	0	0
5	Leora Sissens	0	1437.47	4.48	34	0	1	0	1
6	Rafferty Broske	1	2439.48	4.2	30	0	1	1	1
7	Leta Haydon	0	283.64	5.75	44	1	0	0	1
8	Carolann Brooksbie	1	2021.98	4.58	43	0	1	1	0
9	Conrad Addey	1	691.35	5.15	49	0	1	0	1
10	Deeann Gadault	0	2058.29	5.92	61	1	1	0	1
11	Miller Camachen	1	1209.36	6.07	23	1	0	0	0

Query executed successfully. LAPTOP-P1HT7PG6 (11.0 RTM) sa (55) db_clientes_frecuentes 00:00:00 330 rows

Con estos resultados ahora el Señor Diego puede definir un plan específico para cada uno de sus clientes, por ejemplo descuentos para los clientes frecuentes y promociones para sus clientes regulares.

A continuación las características de cada grupo obtenidas mediante consultas hechas en ambas bases de datos:

Cantidad de servicios tomados:

```
select sum(Sauna) as Sauna_Cantidad, sum(Masaje) as Masaje_Cantidad,
sum(Hidro) as Hidro_Cantidad, sum(Yoga) as Yoga_Cantidad from frecuente
```

	Sauna_Cantidad	Masaje_Cantidad	Hidro_Cantidad	Yoga_Cantidad
1	142	181	130	176

```
select sum(Sauna) as Sauna_Cantidad, sum(Masaje) as Masaje_Cantidad,
sum(Hidro) as Hidro_Cantidad, sum(Yoga) as Yoga_Cantidad from regular
```

	Sauna_Cantidad	Masaje_Cantidad	Hidro_Cantidad	Yoga_Cantidad
1	142	185	141	185

Como podemos observar tanto los clientes frecuentes como los regulares toman en promedio la misma cantidad de servicios.

Cantidad de hombres y mujeres

```
select count(Sexo) as numero_de_hombres
from frecuente
where Sexo LIKE 1
```

	numero_de_hombres
1	146

```
select count(Sexo) as numero_de_hombres
from regular
where Sexo LIKE 1
```

	numero_de_hombres
1	158

</

En ambas poblaciones se nota que el sexo predominante es el de las mujeres, ya que ,frecuentan más los establecimientos en comparación a los hombres.

Cantidad de ingresos para cada categoría

<pre>select sum(Ingresos) as Ingresos_Totales from frecuente</pre>		
<div>100 % <</div> <div>Results Messages</div> <table> <tr> <th>Ingresos_Totales</th></tr> <tr> <td>1 566460.69</td></tr> </table>	Ingresos_Totales	1 566460.69
Ingresos_Totales		
1 566460.69		
<pre>select sum(Ingresos) as Ingresos_Totales from regular</pre>		
<div>100 % <</div> <div>Results Messages</div> <table> <tr> <th>Ingresos_Totales</th></tr> <tr> <td>1 609379.74</td></tr> </table>	Ingresos_Totales	1 609379.74
Ingresos_Totales		
1 609379.74		

Con respecto a la cantidad de ingresos en ambos grupos se determino que el grupo de los clientes regulares genera más ingresos. **Todo este análisis ayudará al Spa de Diego en su campaña de fidelización para que así elija bien su público al cual estará destinada la campaña.**

Ejercicio 2

Para el análisis de la Floristería, hemos dispuesto de la creación de 4 data Flow uno del país y los otros 3 de los tres departamentos.

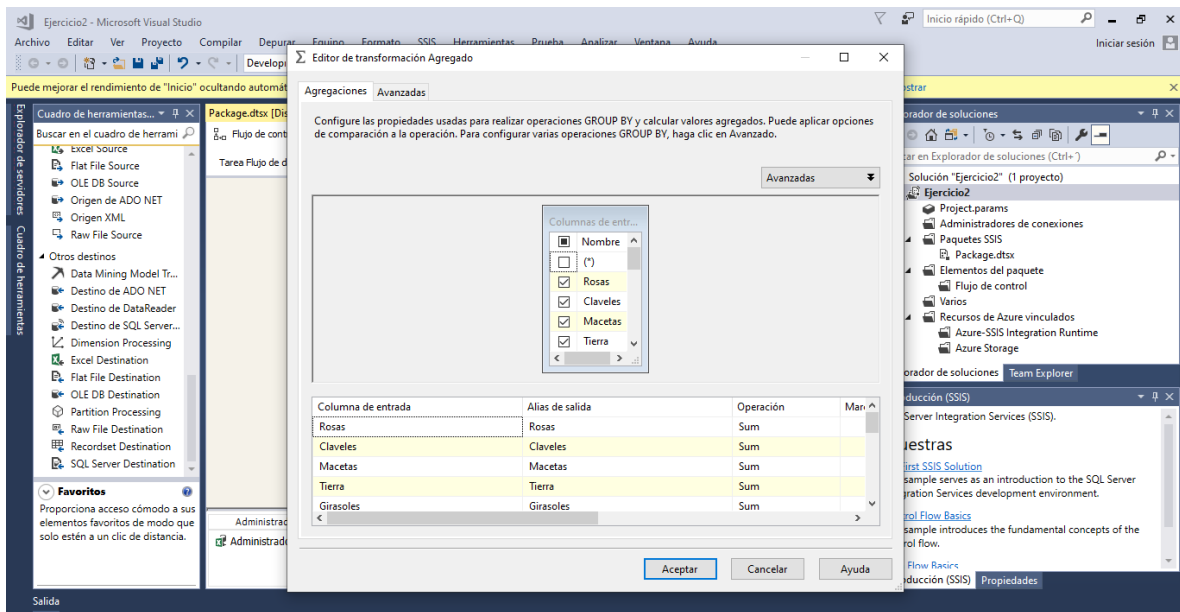
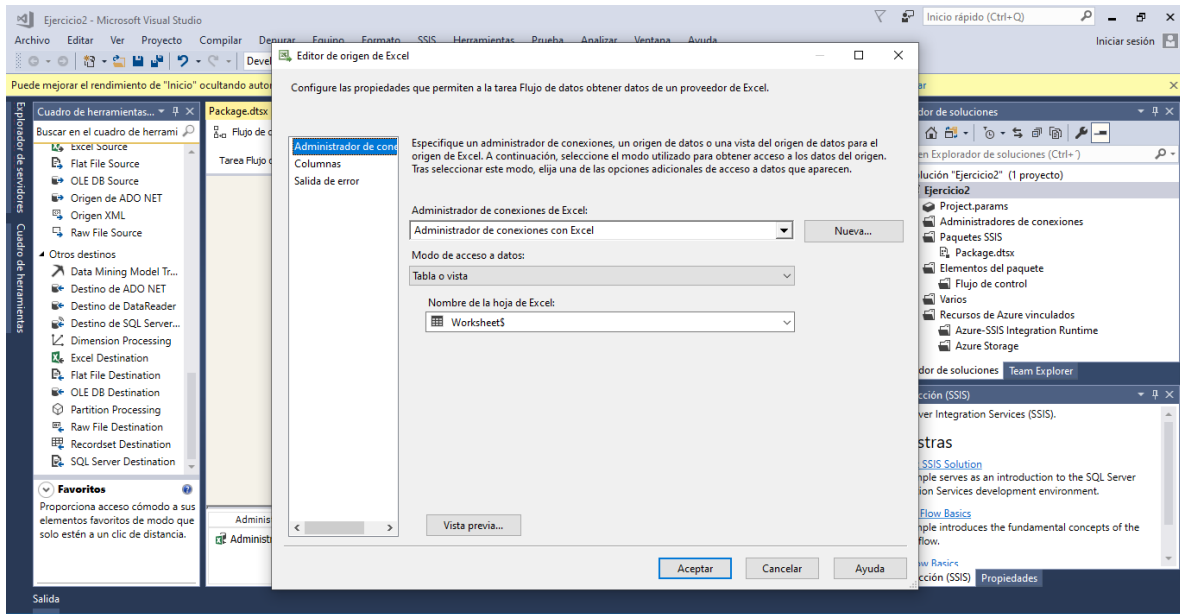


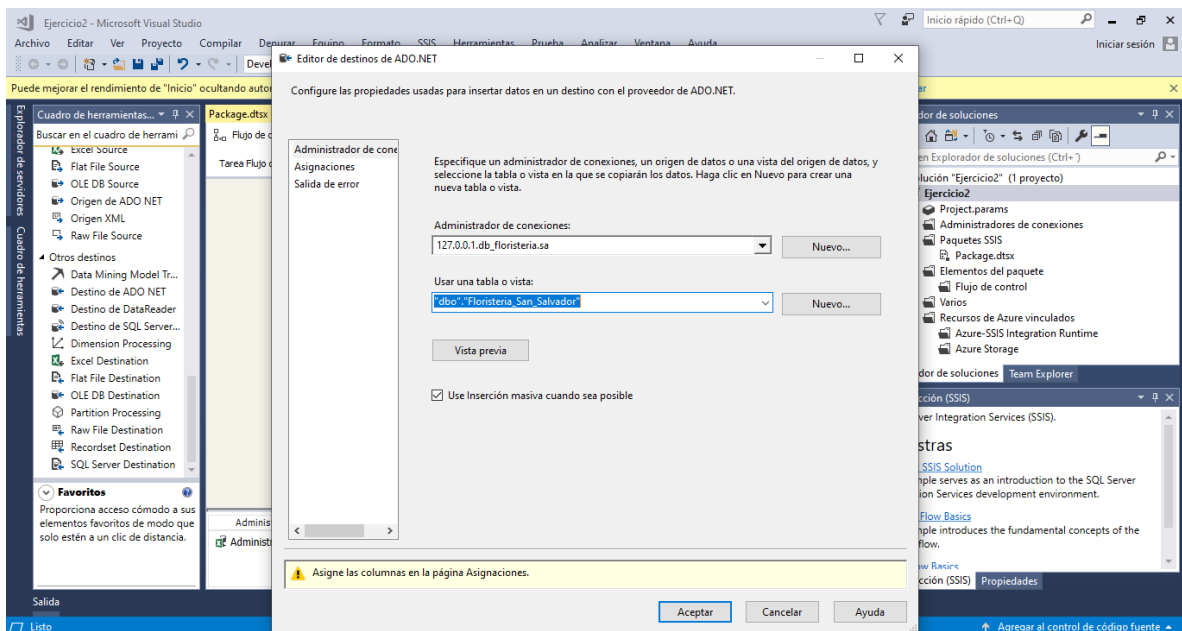
La configuración de los tres data Flow para los departamentos sería de la siguiente manera:





Proseguimos con la configuración general de todos los data flow de cada departamento, desde la fuente hasta el destino:

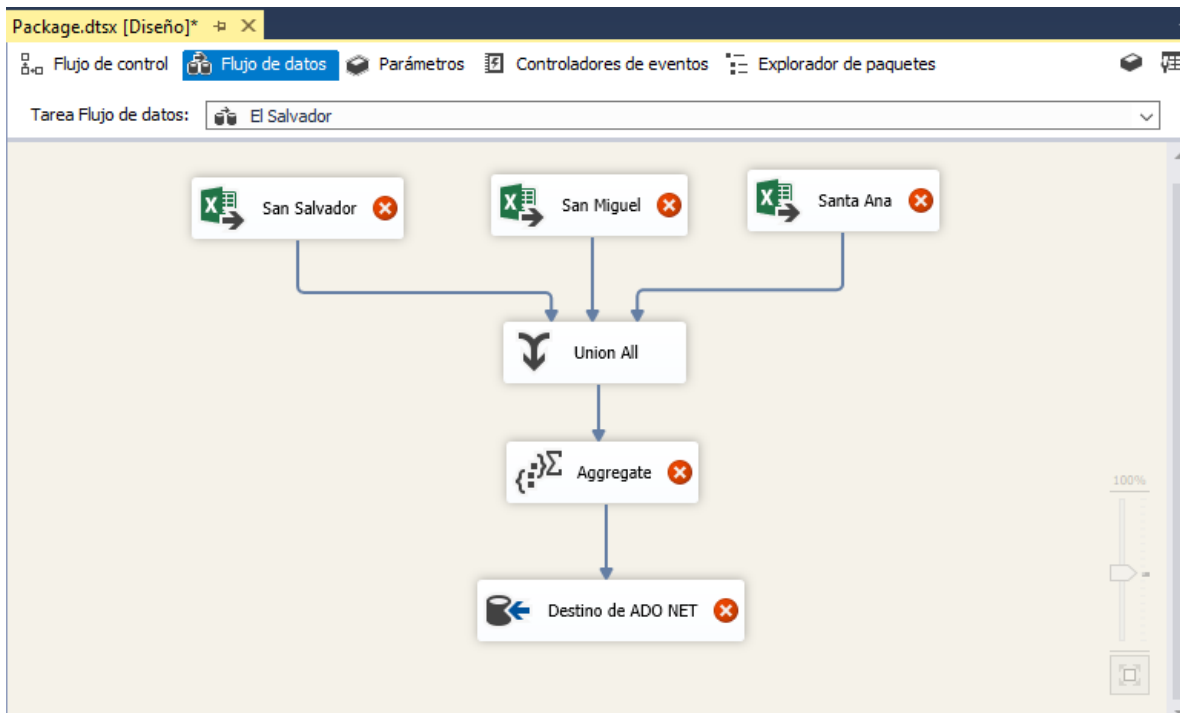




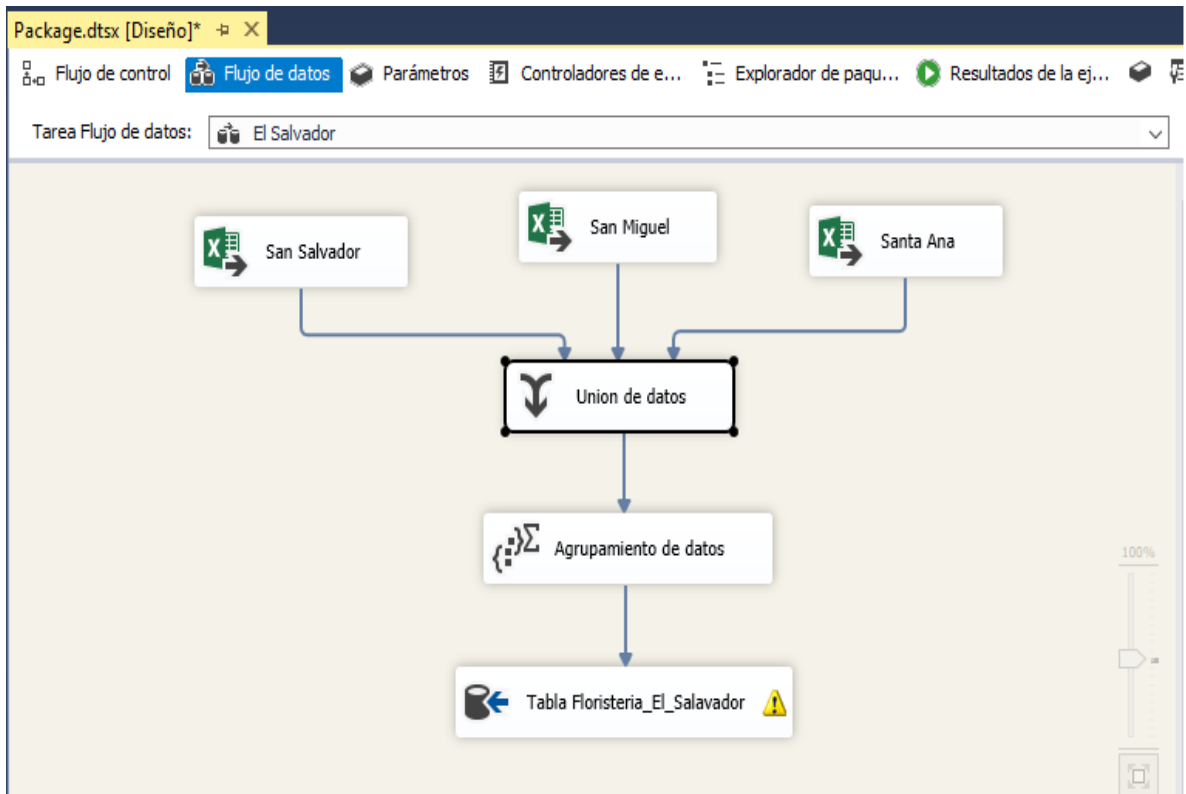
Para los tres quedaría así:

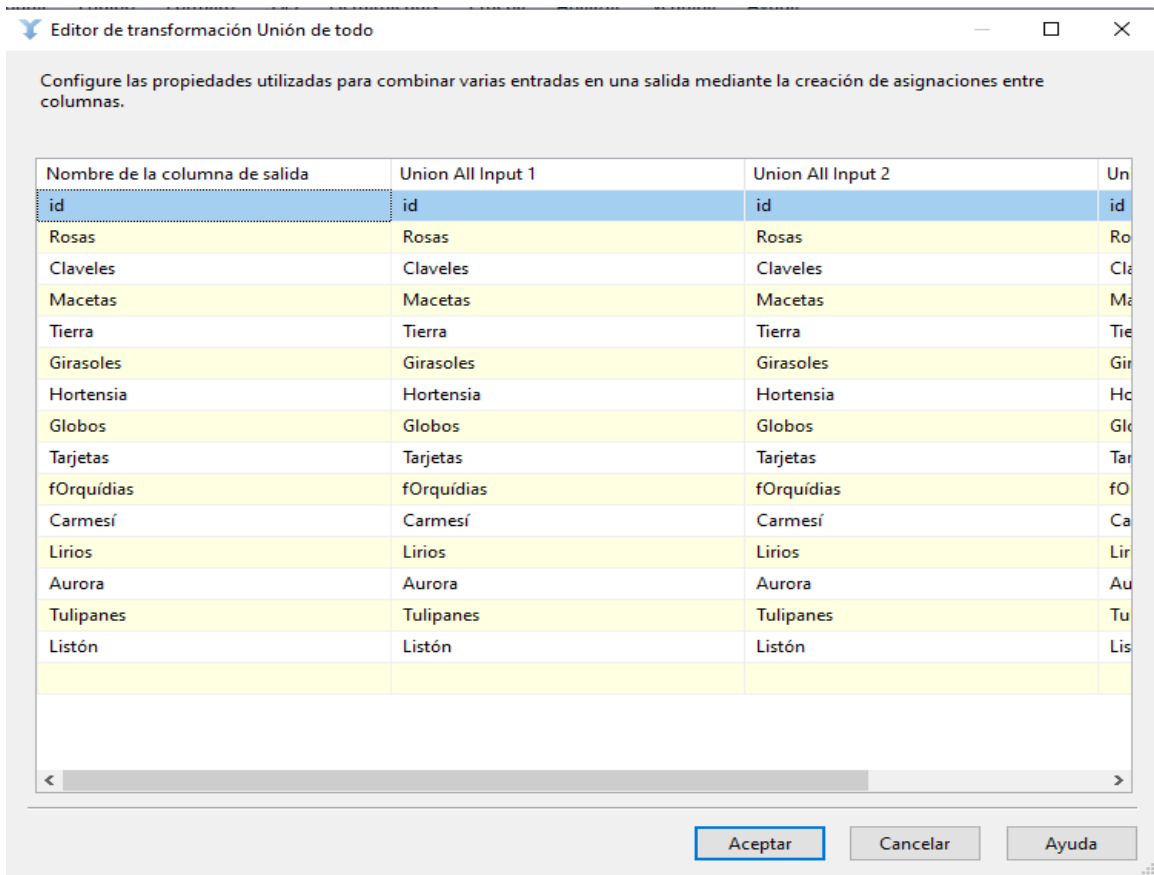


Para la configuración de la data Flow de país, se dispondría de la siguiente manera.:

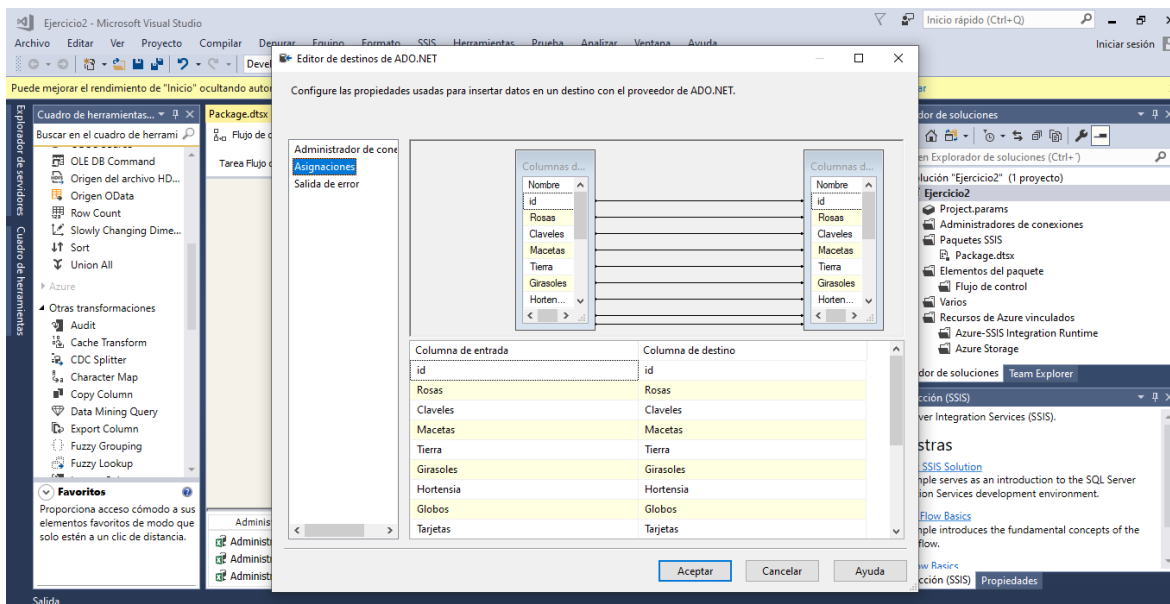


La configuración de todos los elementos sería exactamente como se hizo en los data Flow de los departamentos, con la única diferencia que en este se unen los datos:





Proseguimos con la configuración del destino:

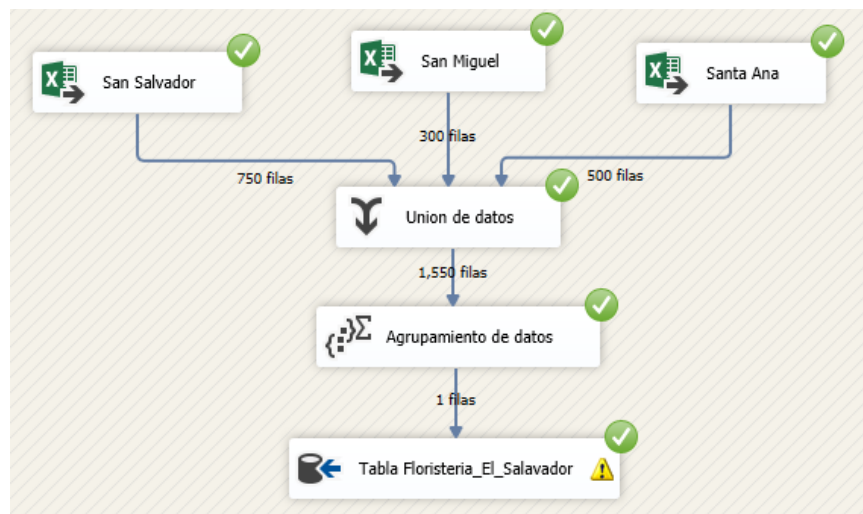
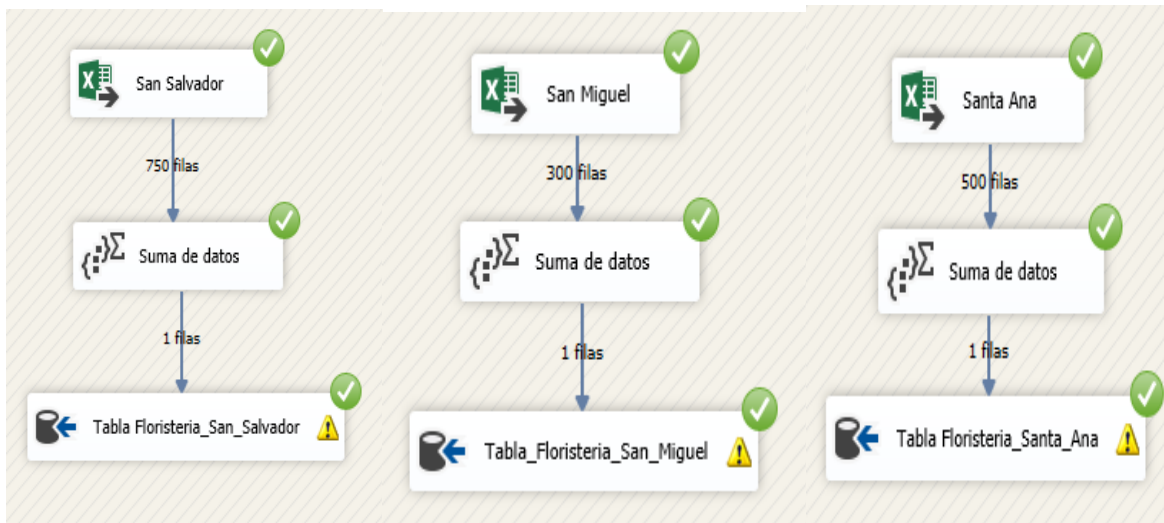


Este sería el resultado después de todas las configuraciones:



Proseguimos con la ejecución del ETL:





Los resultados en la base de datos son:

```
select*from Floristeria_San_Salvador
```

100 %

Results Messages

	id	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquidias	Cemesi	Lirios	Aurora	Tulipanes	Liston
1	750	612	350	392	368	371	374	587	384	380	353	365	384	357	690

Podemos concluir que los productos más destacados para la sucursal de San Salvador son: La rosas, los listones y los globos.

```
select*from Floristeria_San_Miguel
```

100 %

Results Messages

	id	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquidias	Cemesi	Lirios	Aurora	Tulipanes	Liston
1	300	157	137	141	141	150	157	151	143	158	158	160	160	149	149

Para la sucursal de San Miguel concluimos que los productos más destacados son: Aurora, Lirios, Orquidias, Carmesí, Hortensias, y Rosas. Notamos que los rangos no varían mucho. Para esta tienda los productos se venden uniformemente.

```
select*from Floristeria_Santa_Ana
```

100 %

Results

Messages

	id	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquidias	Cemesi	Lirios	Aurora	Tulipanes	Liston
1	500	176	246	245	236	266	243	154	252	259	236	270	260	247	136

Analizando la sucursal de Santa Ana, los productos más destacados son: Lirios, Aurora, y Girasoles. Aunque se han vendido más de 200 unidades la mayoría de los productos. Recomendación hacer una combinación entre los menos vendidos para aumentar la posibilidad de compra.

```
select*from Floristeria_El_Salvador
```

100 %

Results

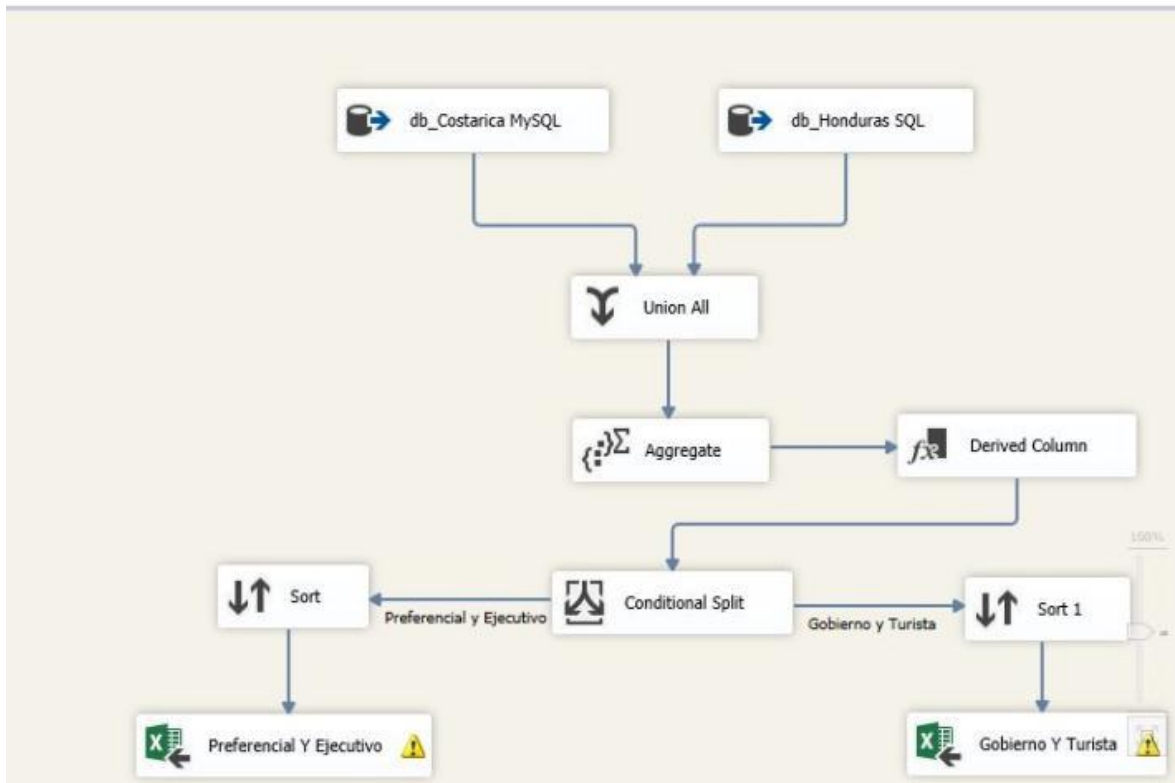
Messages

	id	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquidias	Cemesi	Lirios	Aurora	Tulipanes	Liston
1	1550	945	733	778	745	787	774	892	779	797	747	795	804	753	975

Un análisis de todo el país, en total 1550 clientes, concluimos que los productos más destacados son: Rosas, Listones y los globos. Se recomienda hacer un paquete que combine los tres mejores productos globales. O sea: Las rosas, listones y globos.

Ejercicio 3

Para este ejercicio la composición de los elementos quedaría de la siguiente manera:



La configuración de los orígenes de datos quedaría de la siguiente manera:

Editor de orígenes de ADO.NET

Configure las propiedades usadas por un flujo de datos para obtener datos de un proveedor de ADO.NET.

Administrador de conexiones de ADO.NET: Nueva...

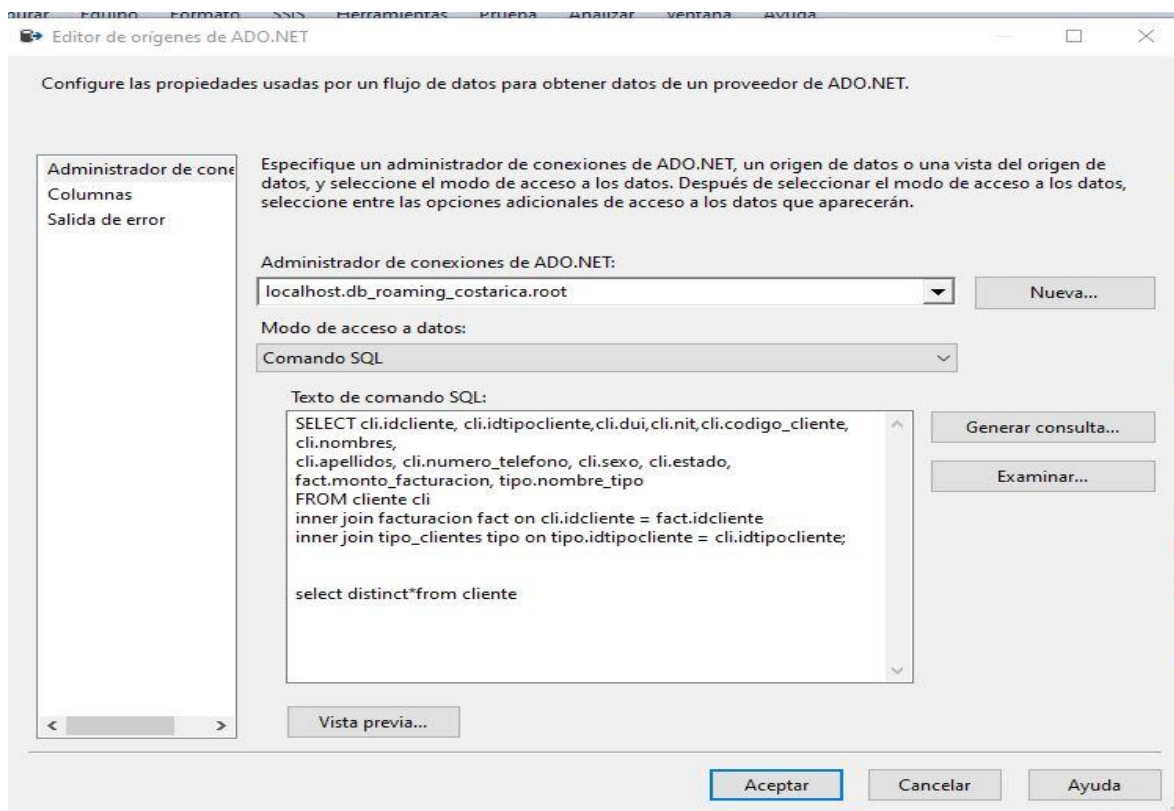
Modo de acceso a datos:

Texto de comando SQL:

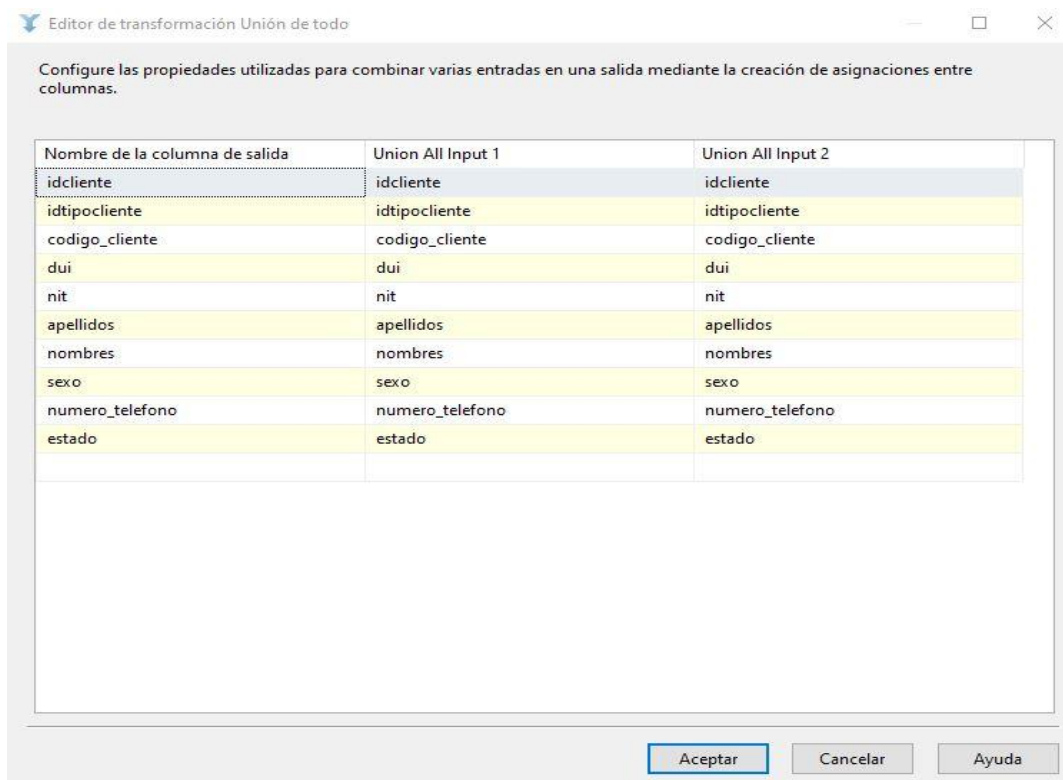
```
SELECT cli.idcliente, cli.idtipocliente, cli.dui, cli.nit, cli.codigo_cliente, cli.nombres, cli.apellidos, cli.numero_telefono, cli.sexo, cli.estado, fact.monto_facturacion, tipo.nombre_tipo FROM cliente cli inner join facturacion fact on cli.idcliente = fact.idcliente inner join tipo_clientes tipo on tipo.idtipocliente = cli.idtipocliente
```

Generar consulta... Examinar... Vista previa...

Aceptar Cancelar Ayuda



Seguimos con la configuración de la unión de todos los campos:



Continuamos con la agrupación de todos los datos

Σ Editor de transformación Agregado

Agregaciones Avanzadas

Configure las propiedades usadas para realizar operaciones GROUP BY y calcular valores agregados. Puede aplicar opciones de comparación a la operación. Para configurar varias operaciones GROUP BY, haga clic en Avanzado.

Avanzadas ▼

Columnas de entrada disponibles

- ☒ Nombre
- ☐ (*)
- ☒ idcliente
- ☒ idtipocliente
- ☒ codigo_cliente
- ☒ dui

Columna de entrada	Alias de salida	Operación	Maneja
idcliente	idcliente	Group by	
idtipocliente	idtipocliente	Group by	
codigo_cliente	codigo_cliente	Group by	
dui	dui	Group by	
nit	nit	Group by	

Aceptar Cancelar Ayuda

Configuramos el formato de los datos con el controlador de las columnas derivadas

Editor de transformación Columna derivada

Especifique las expresiones utilizadas para crear nuevos valores de columna e indique si los valores actualizan las columnas existentes o bien llenan columnas nuevas.

Variables y parámetros

Columnas

Funciones matemáticas

Funciones de cadena

Funciones de fecha y hora

Funciones NULL

Conversiones de tipo

Operadores

Descripción:

Nombre de columna d...	Columna derivada	Expresión	Tipo de datos	Lo
CODIGO_PAIS	<agregar como colum...	SUBSTRING(codigo_cliente,1,2)	Unicode string [DT_WS...	2
DUI_FORMATO	<agregar como colum...	SUBSTRING(dui,1,8) + "-" + SUBSTRING(dui,9,1)	Unicode string [DT_WS...	1
NIT_FORMATO	<agregar como colum...	SUBSTRING(nit,1,4) + "-" + SUBSTRING(nit,5,6) + "-...	Unicode string [DT_WS...	1
NOMBRE_COMPLETO	<agregar como colum...	nombres + "," + apellidos	Unicode string [DT_WS...	5
SEXO_LABEL	<agregar como colum...	sexo == "m" ? "MASCULINO" : "FEMENINO"	Unicode string [DT_WS...	9
TELEFONO_FORMATO	<agregar como colum...	SUBSTRING(numero_telefono,1,4) + "-" + SUBSTRI...	Unicode string [DT_WS...	9
ESTADO_LABEL	<agregar como colum...	estado == "t" ? "ACTIVO" : "INACTIVO"	Unicode string [DT_WS...	8

< >

Configurar la salida de errores... Aceptar Cancelar Ayuda

Seguimos con una división condicional, para enviar a los datos a su base de datos correspondiente

Editor de transformación División condicional

Especifique las condiciones utilizadas para dirigir filas de entrada a salidas específicas. Si una fila de entrada no coincide con ninguna condición, se dirigirá a una salida predeterminada.

Variables y parámetros

Columnas

Funciones matemáticas

Funciones de cadena

Funciones de fecha y hora

Funciones NULL

Conversiones de tipo

Operadores

Descripción:

Orden	Nombre de salida	Condición
1	Preferencial y Ejecutivo	idtipocliente == 1
2	Gobierno y Turista	idtipocliente == 2

Nombre de salida predeterminado: Conditional Split Default Output

Configurar la salida de errores... Aceptar Cancelar Ayuda

Ordenamos los datos para ambos destinos

Editor de transformación Ordenar

Especifique las columnas que se ordenarán y establezca el tipo y el criterio de ordenación. Las columnas no seleccionadas se copiarán sin ninguna modificación.

Columnas de entrada disponibles

Nombre	Paso a través
<input checked="" type="checkbox"/> Nombre	<input type="checkbox"/>
<input type="checkbox"/> idcliente	<input checked="" type="checkbox"/>
<input type="checkbox"/> idtipocliente	<input checked="" type="checkbox"/>
<input type="checkbox"/> codigo_cliente	<input checked="" type="checkbox"/>
<input type="checkbox"/> dui	<input checked="" type="checkbox"/>
<input type="checkbox"/> nit	<input checked="" type="checkbox"/>

Columna de entrada	Alias de salida	Tipo de orden	Criterio de or...	Mar
DUI_FORMATO	DUI_FORMATO	ascendente	1	

☐ Quitar filas con valores de ordenación duplicados Aceptar Cancelar Ayuda

Editor de transformación Ordenar

Especifique las columnas que se ordenarán y establezca el tipo y el criterio de ordenación. Las columnas no seleccionadas se copiarán sin ninguna modificación.

Columnas de entrada disponibles

Nombre	Paso a través
<input checked="" type="checkbox"/> Nombre	<input type="checkbox"/>
<input type="checkbox"/> idcliente	<input checked="" type="checkbox"/>
<input type="checkbox"/> idtipocliente	<input checked="" type="checkbox"/>
<input type="checkbox"/> codigo_cliente	<input checked="" type="checkbox"/>
<input type="checkbox"/> dui	<input checked="" type="checkbox"/>
<input type="checkbox"/> nit	<input checked="" type="checkbox"/>

Columna de entrada	Alias de salida	Tipo de orden	Criterio de or...	Mar
NOMBRE_COMPLETO	NOMBRE_COMPLETO	descendente	1	

☐ Quitar filas con valores de ordenación duplicados Aceptar Cancelar Ayuda

Proseguimos con la configuración del destino de datos y seguidamente la ejecución del ETL:

