


Creación y uso de un flujo de datos (Gen2) en Microsoft Fabric

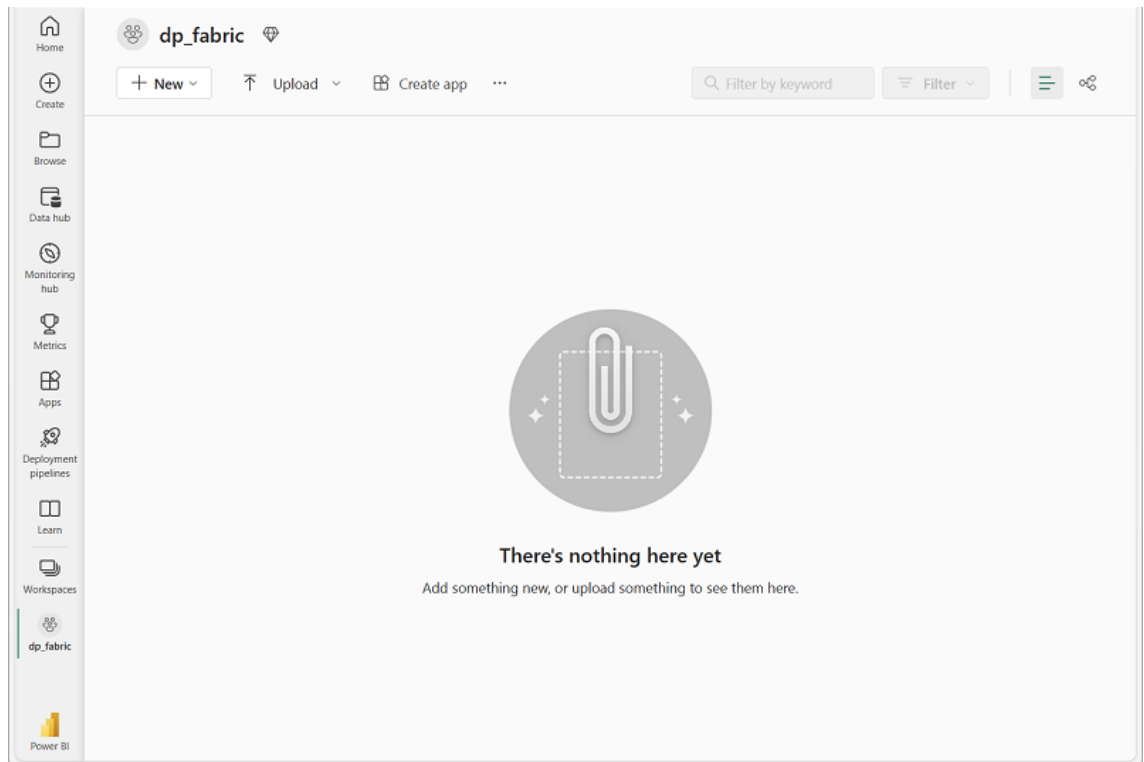
En Microsoft Fabric, los flujos de datos (Gen2) se conectan a varios orígenes de datos y realizan transformaciones en Power Query Online. Luego, se pueden usar en canalizaciones de datos para ingerir datos en un almacén de lago u otro almacén analítico, o para definir un conjunto de datos para un informe de Power BI.

Este laboratorio está diseñado para introducir los distintos elementos de los flujos de datos (Gen2) y no para crear una solución compleja que pueda existir en una empresa.

Creación de un área de trabajo

Antes de trabajar con datos de Fabric, cree un área de trabajo con la evaluación gratuita de Fabric habilitada.

1. En la [página principal de Microsoft Fabric](#), seleccione **Ingeniería de datos de Synapse**.
2. En la barra de menús de la izquierda, seleccione **Áreas de trabajo** (el icono tiene un aspecto similar a .
3. Cree una nueva área de trabajo con el nombre que prefiera y seleccione un modo de licencia que incluya capacidad de Fabric (*Evaluación gratuita, Prémium o Fabric*).
4. Cuando se abra la nueva área de trabajo, debe estar vacía.

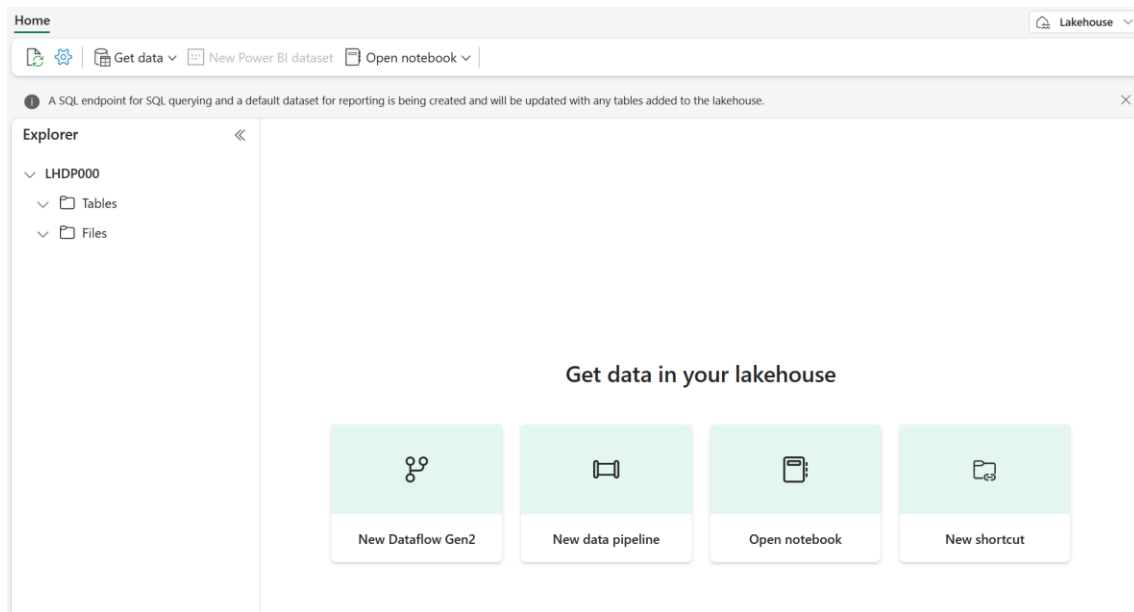


Crear un almacén de lago

Ahora que tiene un área de trabajo, es el momento de crear un almacén de lago de datos en el cual ingerirá los datos.

1. En la página principal de **Ingeniería de datos de Synapse**, cree un nuevo **almacén de lago** con el nombre que prefiera.

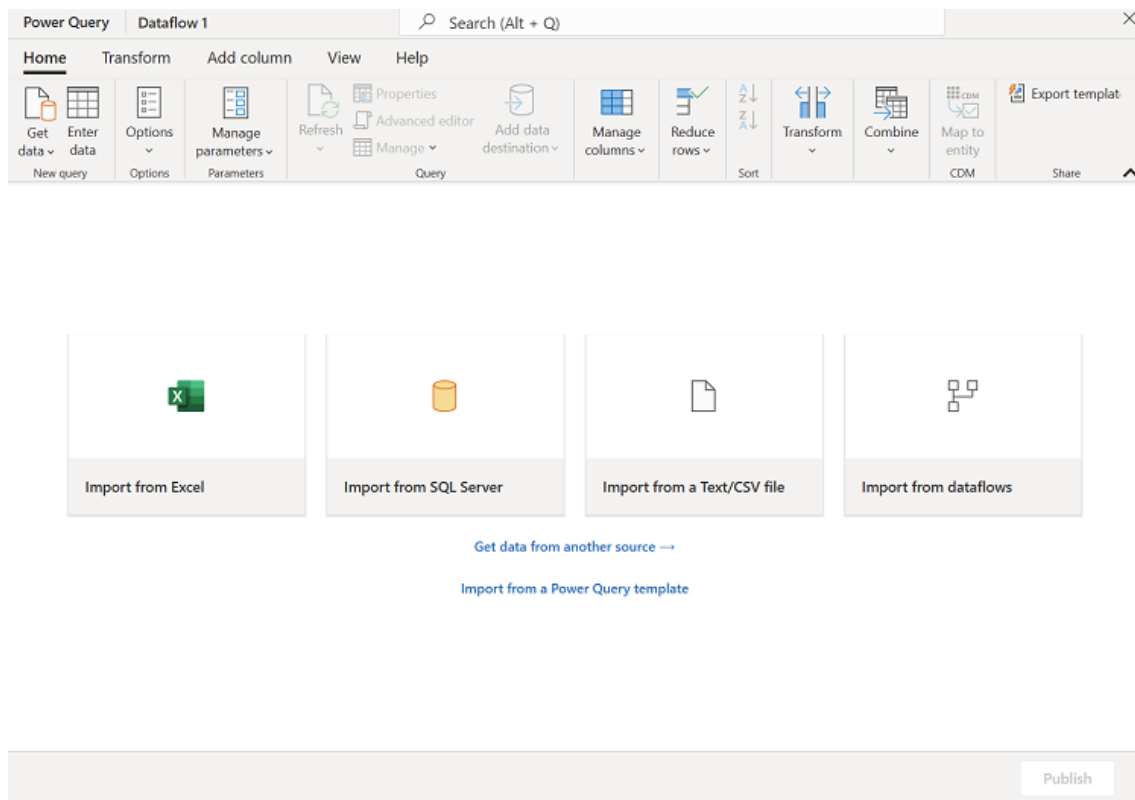
Al cabo de un minuto más o menos, se creará un nuevo almacén de lago vacío.



Creación de un flujo de datos (Gen2) para ingerir datos

Ahora que tiene un almacén de lago, debe ingerir en él algunos datos. Una manera de hacerlo es definir un flujo de datos que encapsula un proceso de *extracción, transformación y carga* (ETL).

1. En la página principal del área de trabajo, seleccione **Nuevo flujo de datos Gen2**. Al cabo de unos segundos, se abre el Editor de Power Query para el nuevo flujo de datos, como se muestra aquí.



1. Seleccione **Importar desde un archivo de texto o CSV** y cree un nuevo origen de datos con la siguiente configuración:
 - **Vínculo al archivo:** *Seleccionado*.
 - **URL o ruta del archivo:** <https://raw.githubusercontent.com/MicrosoftLearning/dp-data/main/orders.csv>.
 - **Conexión:** Crear nueva conexión.
 - **Puerta de enlace de datos:** (ninguna).
 - **Tipo de autenticación:** Anónima.
2. Seleccione **Siguiente** para obtener una vista previa de los datos del archivo y, luego **Crear** para crear el origen de datos. El Editor de Power Query muestra el origen de datos y un conjunto inicial de pasos de consulta para dar formato a los datos, como se muestra aquí:

Home	Transform	Add column	View	Help
Get data	Options	Manage parameters	Refresh	Properties
New query	Options	Parameters	Advanced editor	Manage
		Add data destination	Choose columns	Remove columns
		Manage columns	Remove rows	Keep rows
			Remove rows	Filter rows
			Sort	Transform
			Combine	Map to entity
			CDM	Export template
			Share	

Queries [1]	Table.TransformColumnTypes(*Promoted headers*, {{["SalesOrderID", Int64.Type], ["OrderDate", type date],
orders	

SalesOrderID	OrderDate	CustomerID	LineItem	ProductID	OrderQty	LineItemTotal
1	7/1/2022	29847	1	836	1	356.9
2	7/1/2022	29847	2	822	1	356.9
3	7/1/2022	30072	1	907	1	63.9
4	7/1/2022	30113	1	905	4	873.82
5	7/1/2022	30113	2	983	2	923.39
6	7/1/2022	30113	3	988	6	406.79
7	7/1/2022	30113	4	748	2	1637.4
8	7/1/2022	30113	5	990	1	323.99
9	7/1/2022	30113	6	926	1	149.87
10	7/1/2022	30113	7	743	1	809.76
11	7/1/2022	30113	8	782	4	5507.98
12	7/1/2022	30113	9	918	2	316.86
13	7/1/2022	30113	10	780	4	5567.98
14	7/1/2022	30113	11	937	1	48.59
15	7/1/2022	30113	12	867	6	251.96
16	7/1/2022	30113	13	985	1	67.8
17	7/1/2022	30113	14	989	2	647.99
18	7/1/2022	30113	15	991	3	971.98
19	7/1/2022	30113	16	992	1	323.99
20	7/1/2022	30113	17	993	2	647.99
21	7/1/2022	30113	18	984	2	135.6
22	7/1/2022	30113	19	986	3	203.4
23	7/1/2022	30113	20	987	3	203.4
24	7/1/2022	30113	21	981	2	923.39
25	7/1/2022	30113	22	982	3	1385.08
26	7/1/2022	30113	23	782	6	4584.07

Completed (1.49 s) Columns: 7 Rows: 26+

1. En la cinta de la barra de herramientas, seleccione la pestaña **Agregar columna**. A continuación, seleccione **Columna personalizada** y cree una nueva columna.
2. Establezca el *Nuevo nombre de columna* en **MonthNo**, establezca el *Tipo de datos* en **Número entero** y agregue la siguiente fórmula: **Date.Month([OrderDate])**, tal y como se muestra aquí:

Custom column ?

Add a column that is computed from other columns or values.

New column name * <input type="text" value="MonthNo"/>	Data type <input type="text" value="Whole number"/>
Custom column formula * ⓘ <input type="text" value="= Date.Month([OrderDate])"/>	Available column(s) <div> SalesOrderID OrderDate CustomerID LineItem ProductID OrderQty LineItemTotal </div> <input type="button" value="Insert column"/>

[Learn more about Power Query formulas](#)

OK

Cancel

- Query

Home

Transform

Add column

View

Help

Column examples

columns

Invoke custom function

Conditional columns

Rank columns

Duplicate columns

Format

ABC Extract

Cluster values

Statistics

Standard Scientific

Information

Trigonometry

Rounding

Information

Date

Time

Duration

From text

From number

Date and time column

×

fx

Table.TransformColumnTypes(Table.AddColumn(#"Changed column type", "MonthNo", each Date.Month([OrderDate])), {"MonthNo", Int64.Type}))

Queries [1]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	SalesOrderID	OrderDate	CustomerID	LineItem	ProductID	OrderQty	LineItemTotal	MonthNo														
1	71774	6/1/2022	29847	1	836	1	326.9	6														
2	71774	6/1/2022	29847	2	822	1	326.9	6														
3	71776	6/1/2022	30072	1	907	1	63.9	6														
4	71780	6/1/2022	30113	1	905	4	873.82	6														
5	71780	6/1/2022	30113	2	983	2	923.39	6														
6	71780	6/1/2022	30113	3	988	6	406.79	6														
7	71780	6/1/2022	30113	4	748	2	1637.4	6														
8	71780	6/1/2022	30113	5	990	1	323.99	6														
9	71780	6/1/2022	30113	6	936	1	148.87	6														
10	71780	6/1/2022	30113	7	743	1	809.76	6														
11	71780	6/1/2022	30113	8	782	4	5507.98	6														
12	71780	6/1/2022	30113	9	918	2	316.86	6														
13	71780	6/1/2022	30113	10	780	4	5567.88	6														
14	71780	6/1/2022	30113	11	937	1	48.59	6														
15	71780	6/1/2022	30113	12	867	6	251.96	6														
16	71780	6/1/2022	30113	13	985	1	67.8	6														
17	71780	6/1/2022	30113	14	989	2	647.99	6														
18	71780	6/1/2022	30113	15	991	3	971.88	6														
19	71780	6/1/2022	30113	16	992	1	323.99	6														
20	71780	6/1/2022	30113	17	993	2	647.99	6														
21	71780	6/1/2022	30113	18	984	2	135.6	6														
22	71780	6/1/2022	30113	19	986	3	203.4	6														

Query settings

Properties

Name

orders

Entity type

Custom

Applied steps

Source

Promoted h...

abc: Changed co...

Added cust...

Data destination

No data destination

Step


Publish

Completed (0.97 s) Columns: 8 Rows: 99+


- ## Adición de un destino de datos al flujo de datos

1. En la cinta de opciones de la barra de herramientas, seleccione la pestaña **Inicio**. A continuación, en el menú desplegable **Agregar destino de datos**, seleccione **Almacén de lago**.
2. En el cuadro de diálogo **Conectarse al destino de datos**, edite la conexión e inicie sesión con su cuenta organizativa de Power BI para establecer la identidad que usa el flujo de datos para acceder al almacén de lago.

Connect to data destination



Connection credentials

Connection
Lakehouse (none) 

Connection name
Connection

Data gateway
(none)

Authentication kind
Organizational account

You are not signed in. Please sign in.

Sign in


Cancel

Next

1. Seleccione **Siguiente** y, en la lista de áreas de trabajo disponibles, busque el área de trabajo y seleccione el almacén de lago que creó al principio de este ejercicio. A continuación, especifique una nueva tabla llamada **orders**:

Choose destination target

☒ New table ☐ Existing table

Display options 

dp-000

dp-123

dp-xxx [1]

dpxxx_lakehouse


 A new table will be created in database dpxxx_lakehouse

Table name *
orders

Back

order

Cancel

Next

2. En la página **Elegir configuración de destino**, seleccione **Anexar** y, luego, **Guardar configuración**.

Choose destination settings

Update method



Column mapping

<input checked="" type="checkbox"/>	Source	Source type	Destination	Destination type
<input checked="" type="checkbox"/>	SalesOrderID	1 ² ₃ Whole number	SalesOrderID	Whole number
<input checked="" type="checkbox"/>	OrderDate	Date	OrderDate	Date
<input checked="" type="checkbox"/>	CustomerID	1 ² ₃ Whole number	CustomerID	Whole number
<input checked="" type="checkbox"/>	LineItem	1 ² ₃ Whole number	LineItem	Whole number
<input checked="" type="checkbox"/>	ProductID	1 ² ₃ Whole number	ProductID	Whole number
<input checked="" type="checkbox"/>	OrderQty	1 ² ₃ Whole number	OrderQty	Whole number
<input checked="" type="checkbox"/>	LineItemTotal	1.2 Decimal number	LineItemTotal	Decimal number
<input checked="" type="checkbox"/>	MonthNo	1 ² ₃ Whole number	MonthNo	Whole number

Back

Cancel

Save settings

- En la barra de menús, abra **Vista** y seleccione **Vista de diagrama**. Tenga en cuenta que el destino **Almacén de lago** se indica con un icono en la consulta en el Editor de Power Query.

- Seleccione **Publicar** para publicar el flujo de datos. A continuación, espere a que se cree el flujo de datos **Dataflow 1** en el área de trabajo.

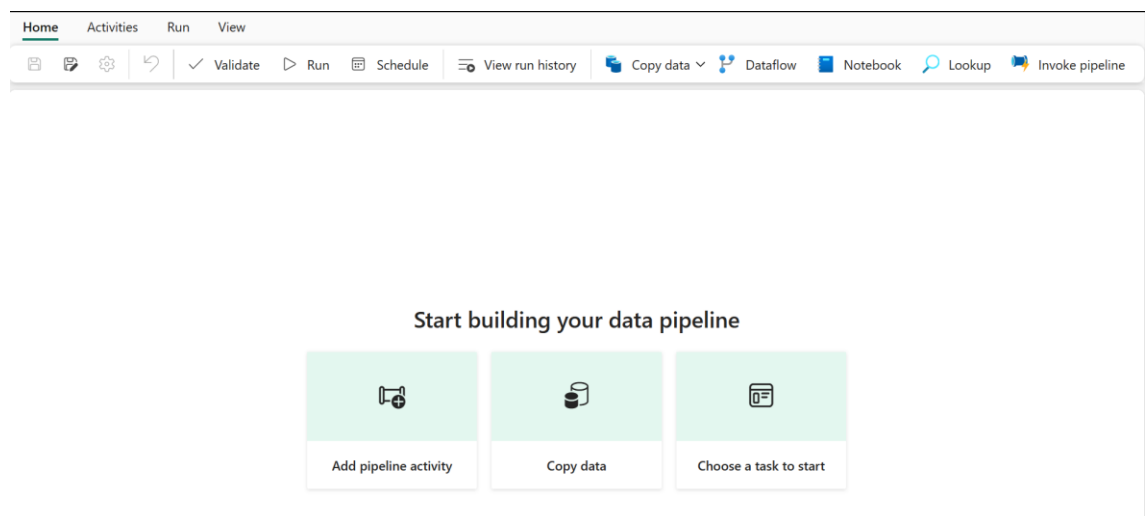
5. Una vez publicado, puede hacer clic con el botón derecho en el flujo de datos del área de trabajo, seleccionar **Propiedades** y cambiar el nombre del flujo de datos.

Adición de un flujo de datos a una canalización

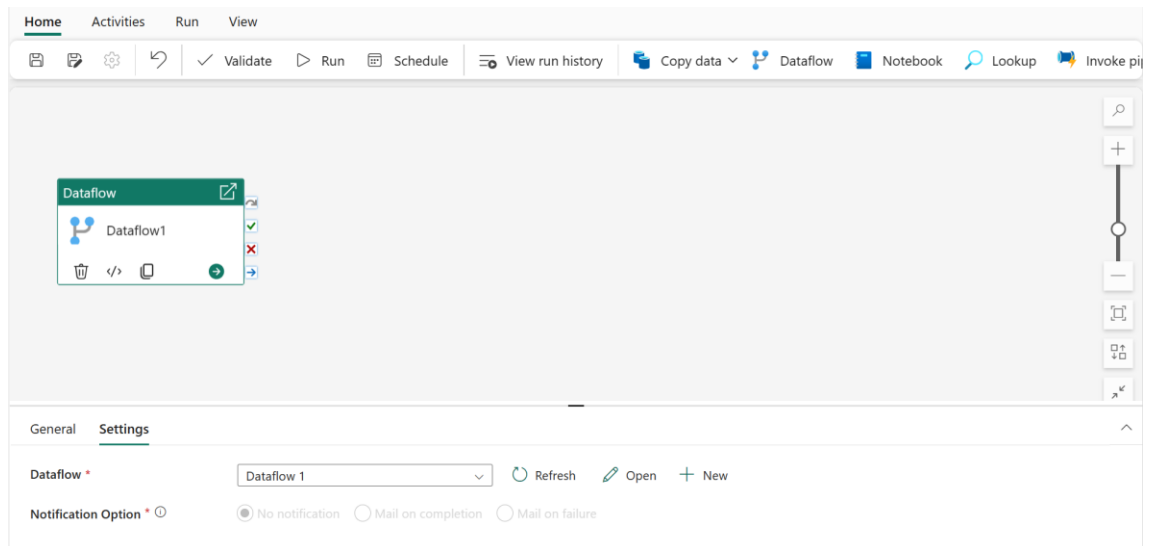
Puede incluir un flujo de datos como actividad en una canalización. Las canalizaciones se usan para orquestar las actividades de ingesta y procesamiento de datos, lo que permite combinar flujos de datos con otros tipos de operaciones en un único proceso programado. Se pueden crear canalizaciones en unas cuantas experiencias diferentes, incluida la experiencia Data Factory.



1. En el área de trabajo habilitada para Fabric, asegúrese de que todavía está en la experiencia **Ingeniería de datos**. Seleccione **Nueva, Canalización de datos** y, cuando se le solicite, cree una canalización llamada **Cargar datos**.

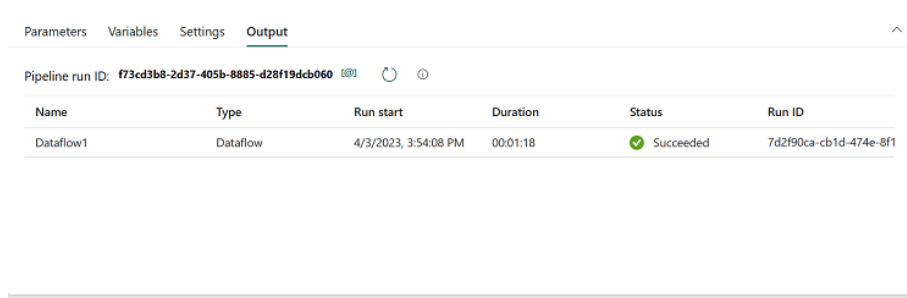
Se abre el editor de canalizaciones.



2. Seleccione **Agregar actividad de canalización** y agregue una actividad **Flujo de datos** a la canalización.
3. Con la nueva actividad **Dataflow1** seleccionada, en la pestaña **Configuración**, en la lista desplegable **Flujo de datos**, seleccione **Dataflow 1** (el flujo de datos que creó anteriormente).



4. En la pestaña **Inicio**, guarde la canalización con el icono  (*Guardar*).
5. Use el botón  **Ejecutar** para ejecutar la canalización y espere a que se complete. Esto puede tardar unos minutos.



6. En la barra de menús del borde izquierdo, seleccione su almacén de lago.
7. En el menú **...** de **Tablas**, seleccione **Actualizar**. A continuación, expanda **Tablas** y seleccione la tabla **orders**, que ha creado el flujo de datos.

