



Etapa 2 Laboratorio 1 - TechGear - Paso a paso (README)

Integrantes: Vanina Correa
Alan Donoso
Xavier Muñoz
Catalina Olivares
John Serrano
Curso: Arquitectura de Datos
Profesora: Valentina Tombolini

9 de Agosto de 2024

Tabla de contenidos

1. Bases de datos y Fakers	1
2. Microsoft Azure	5
3. Power BI	22

1. Bases de datos y Fakers

1. Instalar algún gestor de base de datos de PostgreSQL, MySQL y SQL Server. Para este caso se utiliza PgAdmin 4 con la versión 18 de postgres descargado de <https://www.pgadmin.org/download/>, phpMyAdmin 5.2.1 con la versión 8.0 de MySQL adquirido de <https://www.phpmyadmin.net/> y SSMS 20.2 para SQL Server obtenido de <https://learn.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver15>.
2. En pgAdmin u otro gestor de base de datos de PostgreSQL es necesario conectar con los servidores de las bases de datos: **cliente** y **logistica**. Se deben usar las credenciales:
Para cliente:

```
bd-cliente
host: bd-cliente.postgres.database.azure.com
user: cliente
pass: Arquí1234!
```

Para logística:

```
bd-logistica
host: bd-logistica.postgres.database.azure.com
user: logistica
pass: Arquí1234!
```

3. En PHPMyAdmin u otro gestor de base de datos de MySQL, es necesario conectar con los servidores de las bases de datos: **inventario** y **rrhh**. Se deben usar las credenciales:
Para inventario:

```
bd-inventario
host: bd-inventario.mysql.database.azure.com
```

user: inventario

pass: Arquii234!

Para rrhh:

```
bd-rrhh  
host: bd-rrhh.mysql.database.azure.com  
user: rrhh  
pass: Arquí1234!
```

4. En SQL Server Management Studio u otro gestor de base de datos de SQL Server, es necesario conectar con los servidores de la base de datos **productos**. Se deben usar las credenciales:

```
host: bd-productos.database.windows.net  
usuario: producto  
pass: Arquí1234!
```

5. Luego, es necesario ejecutar los scripts .sql para crear las tablas asociadas a las bases de datos y los scripts .py para poblar las tablas creadas con fakers (datos falsos).
 - a) Para la base de datos cliente, importar o ejecutar el script cliente.sql, el cual genera las tablas **Cliente, Boleta, Detalle_Boleta, Producto y Envio**. Luego, ejecutar el script cliente.py para poblar las tablas creadas anteriormente con fakers.
 - b) Para la base de datos inventario, importar o ejecutar el script inventario.sql, el cual genera las tablas **Producto, Stock_Sucursal, Sucursal, Departamento y Sucursal_Departamento**. Luego, ejecutar el script inventario.py para poblar las tablas creadas anteriormente con fakers.
 - c) Para la base de datos logistica, importar o ejecutar el script logistica.sql, el cual genera las tablas **Producto, Producto_Proveedor, Proveedor, Sucursal, Stock_Sucursal, Departamento y Sucursal_Departamento**. Luego, ejecutar el script logistica.py para poblar las tablas creadas anteriormente con fakers.

- d)* Para la base de datos producto importar o ejecutar el script `producto.sql`, el cual genera las tablas **Proveedor**, **Producto**, **Sucursal** y **Stock_Sucursal**. Luego, ejecutar el script `producto.py` para poblar las tablas creadas anteriormente con fakers.
- e)* Para la base de datos rrhh importar o ejecutar el script `rrhh.sql`, el cual genera las tablas **Sucursal**, **Departamento**, **Sucursal_Departamento** y **Trabajador**. Luego, ejecutar el script `rrhh.py` para poblar las tablas creadas anteriormente con fakers.

2. Microsoft Azure

1. Tras iniciar sesión en Microsoft Azure, se crea un grupo de recursos de la Figura 1 desde Marketplace. Se debe nombrar y seleccionar la zona geográfica más cercana, en la Figura 2 se muestran los detalles del grupo de recursos.



Figura 1: Grupo de recursos.

Crear un grupo de recursos ...

✓ Validación superada.

Datos básicos

Etiquetas

Revisar y crear

Datos básicos

Suscripción

Grupo de recursos

Región

Etiquetas

Ninguno

Azure for Students

teachgear-arqui

East US

Figura 2: Configuración del grupo de recursos.

2. Luego, se crean las bases de datos anteriormente mencionadas desde el grupo de recursos, con sus credenciales respectivas. Para la configuración de seguridad, se otorga el permiso de todas IPs públicas y los servicios dentro de su grupo de recursos.

a) Para las bases de datos cliente y logística de PostgreSQL, se utiliza el recurso de la Figura 3 y las configuraciones se muestran en las Figuras 4 y 5, respectivamente.

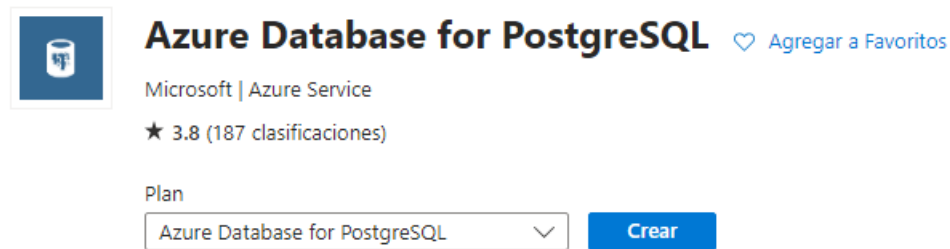


Figura 3: Recurso para las bases de datos de PostgreSQL.

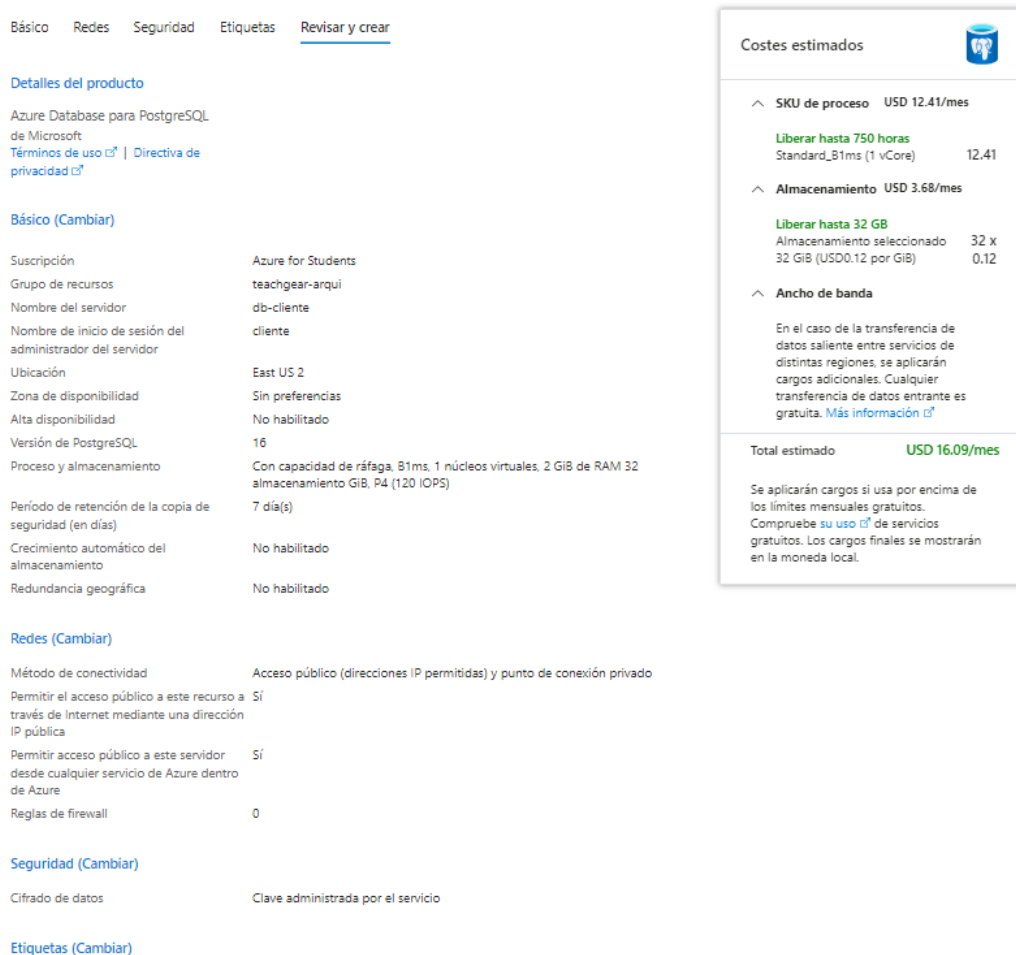


Figura 4: Configuración de la base de datos de cliente.

Básico (Cambiar)

Suscripción	Azure for Students
Grupo de recursos	teachgear-arqui
Nombre del servidor	logistica-bd
Nombre de inicio de sesión del administrador del servidor	logistica
Ubicación	East US 2
Zona de disponibilidad	Sin preferencias
Alta disponibilidad	No habilitado
Versión de PostgreSQL	16
Proceso y almacenamiento	Con capacidad de ráfaga, B1ms, 1 núcleos virtuales, 2 GiB de RAM 32 almacenamiento GiB, P4 (120 IOPS)
Período de retención de la copia de seguridad (en días)	7 día(s)
Crecimiento automático del almacenamiento	No habilitado
Redundancia geográfica	No habilitado

Redes (Cambiar)

Método de conectividad	Acceso público (direcciones IP permitidas) y punto de conexión privado
Permitir el acceso público a este recurso a través de Internet mediante una dirección IP pública	Sí
Permitir acceso público a este servidor desde cualquier servicio de Azure dentro de Azure	Sí
Reglas de firewall	0

Seguridad (Cambiar)

Cifrado de datos	Clave administrada por el servicio
------------------	------------------------------------

Etiquetas (Cambiar)

Costes estimados

^ **SKU de proceso** USD 12.41/mes

Liberar hasta 750 horas

Standard_B1ms (1 vCore) 12.41

^ **Almacenamiento** USD 3.68/mes

Liberar hasta 32 GB

Almacenamiento seleccionado 32 x 32 GiB (USD0.12 por GiB) 0.12

^ **Ancho de banda**

En el caso de la transferencia de datos saliente entre servicios de distintas regiones, se aplicarán cargos adicionales. Cualquier transferencia de datos entrante es gratuita. [Más información](#)

Total estimado **USD 16.09/mes**

Se aplicarán cargos si usa por encima de los límites mensuales gratuitos. Compruebe [su uso](#) de servicios gratuitos. Los cargos finales se mostrarán en la moneda local.

Figura 5: Configuración de la base de datos de logística.

- b) Para las bases de datos rrhh e inventario de MySQL, se utiliza el recurso de la Figura 6 y las configuraciones se muestran en las Figuras 7 y 8, respectivamente.

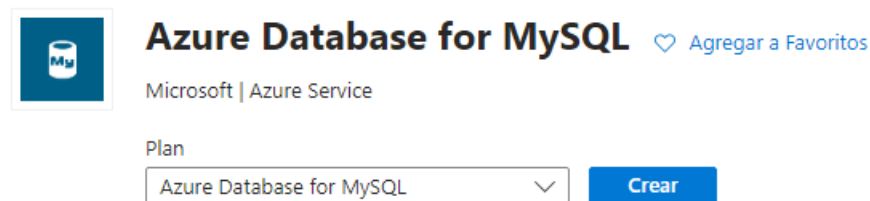


Figura 6: Recurso para las bases de datos de MySQL.

Básico (Cambiar)

Suscripción	Azure for Students
Grupo de recursos	teachgear-arqui
Nombre del servidor	db-rrhh
Nombre de inicio de sesión del administrador del servidor	rrhh
Ubicación	East US
Zona de disponibilidad	Sin preferencias
Alta disponibilidad	No habilitado
Versión de MySQL	8.0
Proceso y almacenamiento	Con capacidad de ráfaga, 81ms, 1 núcleos virtuales, 2 GiB de RAM, 20 almacenamiento, IOPS de escalado automático
Período de retención de la copia de seguridad (en días)	7 día(s)
Registros acelerados	No habilitado
Crecimiento automático del almacenamiento	Habilitado
Redundancia geográfica	No habilitado

Redes (Cambiar)

Método de conectividad	Acceso público (direcciones IP permitidas) y punto de conexión privado
Permitir el acceso público a este recurso a través de Internet mediante una dirección IP pública	Sí
Permitir acceso público a este servidor desde cualquier servicio de Azure dentro de Azure	Sí
Reglas de firewall	0
SSL/TLS	Se aplica SSL y la versión de TLS es la 1.2. Esto se puede cambiar después de crear el servidor. Más información

Seguridad (Cambiar)

Cifrado de datos	Clave administrada por el servicio
------------------	------------------------------------

Etiquetas (Cambiar)

Costes estimados

My

SKU de proceso

USD 12.41/mes

Liberar hasta 750 horas

Standard_81ms (1 vCore)

12.41

Almacenamiento

USD 2.30/mes

Liberar hasta 32 GB

Almacenamiento seleccionado

20 x

20 GiB (USD0.12 por GiB)

0.12

IOPS de escalado automático

El autoescalado de IOPS se factura por uso en incrementos de un millón de solicitudes. [Más información](#)

Retención de la copia de seguridad

La retención de copias de seguridad se factura en función del almacenamiento adicional usado para conservar las copias de seguridad. [Más información](#)

Ancho de banda

En el caso de la transferencia de datos saliente entre servicios de distintas regiones, se aplicarán cargos adicionales. Cualquier transferencia de datos entrante es gratuita. [Más información](#)

Total estimado

USD 14.71/mes

Se aplicarán cargos si usa por encima de los límites mensuales gratuitos. Compruebe [su uso](#) de servicios gratuitos. Los cargos finales se mostrarán en la moneda local.

Figura 7: Configuración de la base de datos de rrhh.

Básico

Redes

Seguridad

Etiquetas

Revisar y crear

Hora de creación

Tiempo estimado de creación del servidor 5 (en minutos)

Detalles del producto

Azure Database para MySQL de Microsoft

[Términos de uso](#)
[Directiva de privacidad](#)

Básico (Cambiar)

Suscripción

Azure for Students

Grupo de recursos

teachgear-arqui

Nombre del servidor

inventario-bd

Nombre de inicio de sesión del administrador del servidor

inventario

Ubicación

East US 2

Zona de disponibilidad

Sin preferencias

Alta disponibilidad

No habilitado

Versión de MySQL

8.0

Proceso y almacenamiento

Con capacidad de ráfaga, 81ms, 1 núcleos virtuales, 2 GiB de RAM, 20 almacenamiento, IOPS de escalado automático

Período de retención de la copia de seguridad (en días)

7 día(s)

Registros acelerados

No habilitado

Crecimiento automático del almacenamiento

Habilitado

Redundancia geográfica

No habilitado

Redes (Cambiar)

Método de conectividad

Acceso público (direcciones IP permitidas) y punto de conexión privado

Permitir el acceso público a este recurso a través de Internet mediante una dirección IP pública

Sí

Permitir acceso público a este servidor desde cualquier servicio de Azure dentro de Azure

Sí

Reglas de firewall

0

SSL/TLS

Se aplica SSL y la versión de TLS es la 1.2. Esto se puede cambiar después de crear el servidor. [Más información](#)

Seguridad (Cambiar)

Cifrado de datos

Clave administrada por el servicio

Etiquetas (Cambiar)

Costes estimados

SKU de proceso

USD 12.41/mes

Liberar hasta 750 horas

Standard_B1ms (1 vCore)

12.41

Almacenamiento

USD 2.30/mes

Liberar hasta 32 GB

Almacenamiento seleccionado

20 x 20 GiB (USD0.12 por GiB)

0.12

IOPS de escalado automático

El autoescalado de IOPS se factura por uso en incrementos de un millón de solicitudes. [Más información](#)

Retención de la copia de seguridad

La retención de copias de seguridad se factura en función del almacenamiento adicional usado para conservar las copias de seguridad. [Más información](#)

Ancho de banda

En el caso de la transferencia de datos saliente entre servicios de distintas regiones, se aplicarán cargos adicionales. Cualquier transferencia de datos entrante es gratuita. [Más información](#)

Total estimado

USD 14.71/mes

Se aplicarán cargos si usa por encima de los límites mensuales gratuitos. Compruebe [su uso](#) de servicios gratuitos. Los cargos finales se mostrarán en la moneda local.

Figura 8: Configuración de la base de datos de inventario.

- c) Para la base de datos de productos de SQL Server, se utiliza el recurso de la Figura 9 y su configuración se muestra en la Figura 10.

SQL Database

Microsoft

SQL

SQL Database

Microsoft | Azure Service

★ 4.4 (2588 clasificaciones)

Plan

SQL Database

Crear

Figura 9: Recurso para la base de datos de SQL Server.

Crear base de datos SQL

Microsoft

Básico

Suscripción	Azure for Students
Grupo de recursos	teachgear-arqui
Región	eastus2
Nombre de la base de datos	bd-productos
Servidor	bd-productos

Proceso y almacenamiento	Básico: Almacenamiento: 2 GB
Redundancia del almacenamiento de copias de seguridad	Almacenamiento de copias de seguridad con redundancia geográfica

Redes

Permitir que los servicios y recursos de Azure accedan a este servidor	Sí
Punto de conexión privado	Ninguno

Seguridad

Cifrado de datos transparente (nivel de servidor)	Clave administrada por el servicio seleccionada
Clave administrada por el cliente en el nivel de base de datos	No configurado
Identidad administrada asignada por el usuario en el nivel de base de datos	No configurado
Advanced Data Security	Ahora no
Always encrypted with secure enclaves	No configurado
Libro de contabilidad Sql(base de datos)	Deshabilitado
Almacenamiento de resumen	Deshabilitado

[Crear](#)
[< Anterior](#)
[Descargar una plantilla para la automatización](#)

Figura 10: Configuración de la base de datos de productos.

- Se debe crear Factoría de datos (V2) en el grupo de recursos, utilizando el recurso de la Figura 11 con la configuración por defecto.



Figura 11: Recurso para Data Factory.

- Se debe crear una cuenta de almacenamiento en el grupo de recursos, utilizando el recurso de la Figura 12 con la configuración por defecto.

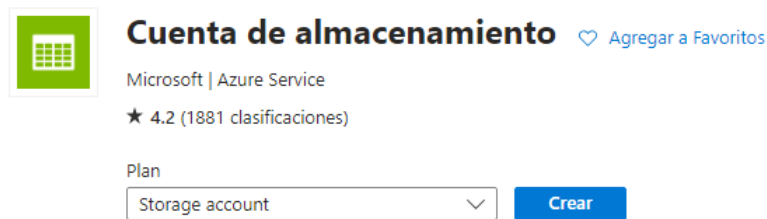


Figura 12: Recurso para Cuenta de Almacenamiento.

- Dentro del recurso techgear, dar click en la cuenta de almacenamiento que se ha creado en el punto 3, en nuestro caso se llama “techgear”. Dentro de esta, en el menú lateral, seleccionar “Almacenamiento de datos”, luego “Contenedores” y dar click en “+ Contenedor”. En “Nombre” indicar el nombre del contenedor y dar click en el botón “crear” como se muestra en la Figura 14. El procedimiento se debe realizar hasta crear las carpetas: “boleta”, “producto”, “proveedor” y “sucursal”.



Recursos		
<div> <div>Reciente</div> <div>Favorito</div> </div>		
Nombre	Tipo	Última consulta
 storageaccounttechgear	Cuenta de almacenamiento	hace 4 minutos
 etl-techgear	Factoría de datos (V2)	hace 3 días

Figura 13: Pestaña de recurso de techgear dentro de Azure.

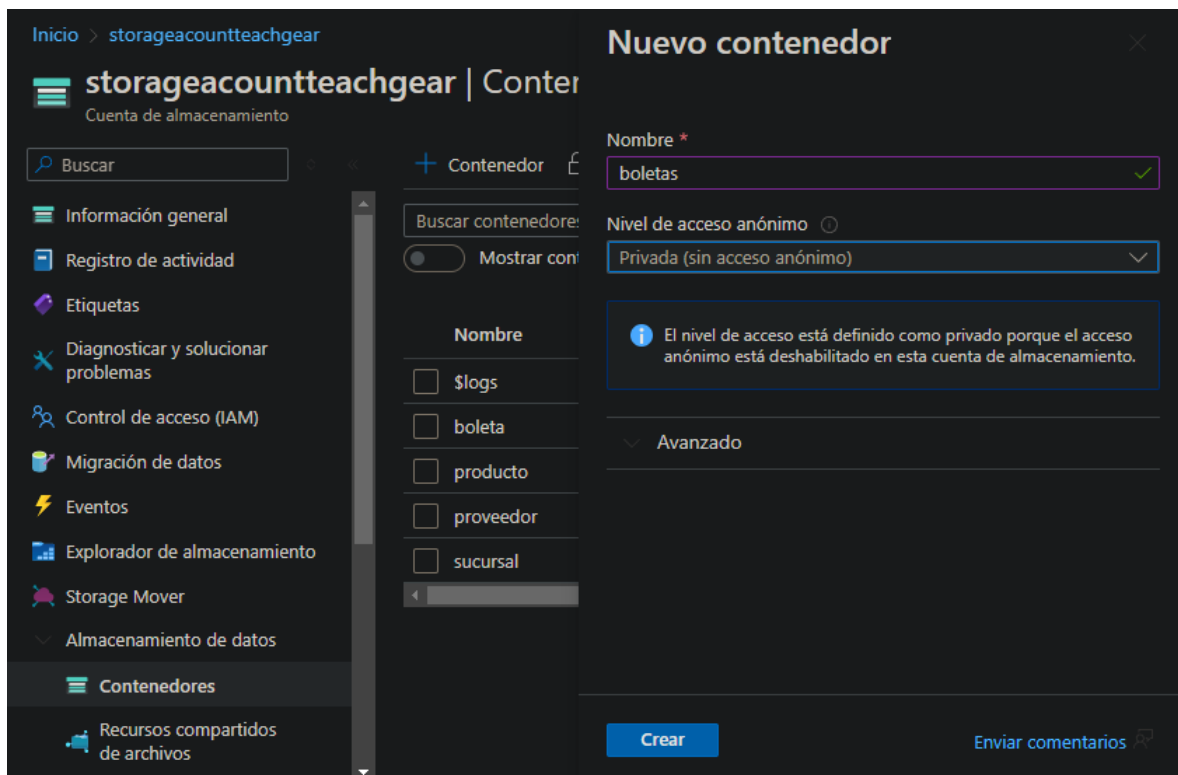


Figura 14: Contenedores en Azure.

6. Luego, se debe regresar al grupo de recursos techgear, ingresar a la factoría de datos creada en el punto 3, en nuestro caso se llama “df-techgear”, y dar click en “Iniciar Studio”.
7. Una vez dentro, ir a “Administrar” (icono de caja de herramientas), “Servicios vinculados”, dar click en “Nuevo”, buscar y seleccionar “Azure Data Lake Storage Gen2”. En “nombre” debe ir un nombre representativo (en nuestro caso es “AzureDataLakeStorage1”), en “Conectar mediante Integración Runtime” seleccionar “AutoResolveIntegrationRuntime” y en “Tipo de autenticación” seleccionar “Clave de cuenta”. En el caso de que se tenga una suscripción de Azure, se debe seleccionar la suscripción de la cuenta que posee la configuración de Azure, en “Nombre de la cuenta de almacenamiento” debe ir la que se ha creado en el punto 4 y luego dar clic en “Crear”.

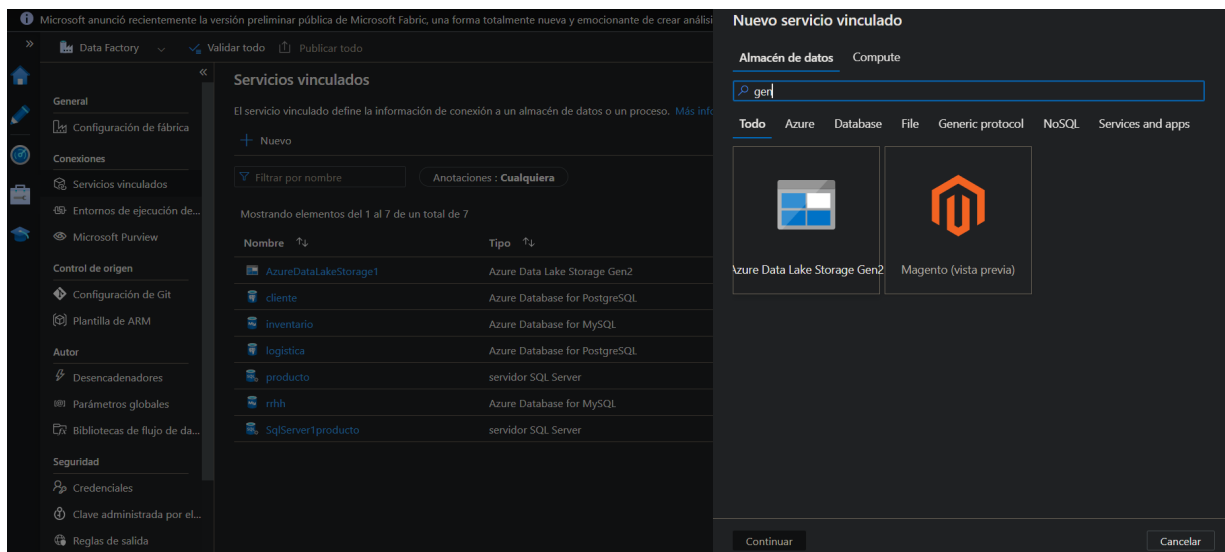


Figura 15: Servicios vinculados.

8. Ahora se deben crear bases de datos para azure para cada BD que se tenga, para ello dar clic en “Nuevo”, buscar y seleccionar “Azure Database for PostgreSQL”, en “Nombre” debe ir el nombre de la base de datos que se ha creado para PostgreSQL y en “Conectar mediante Integración Runtime” seleccionar “AutoResolveIntegrationRuntime”. En el caso de que se tenga una suscripción de Azure, se debe seleccionar la suscripción de la cuenta que posee la configuración de Azure, en “Nombre del servidor” seleccionar la base de datos acorde al nombre ingresado al comienzo, en “Nombre del servidor” seleccionar “postgres”, luego ingresar las credenciales de “nombre de usuario” y “Contraseña” que se utilizaron al momento de crear las bases de datos localmente en el punto 2. Un ejemplo de llenado se muestra en la Figura 16. Esto se debe realizar para las bases de datos de “clientes” y de “logística”.

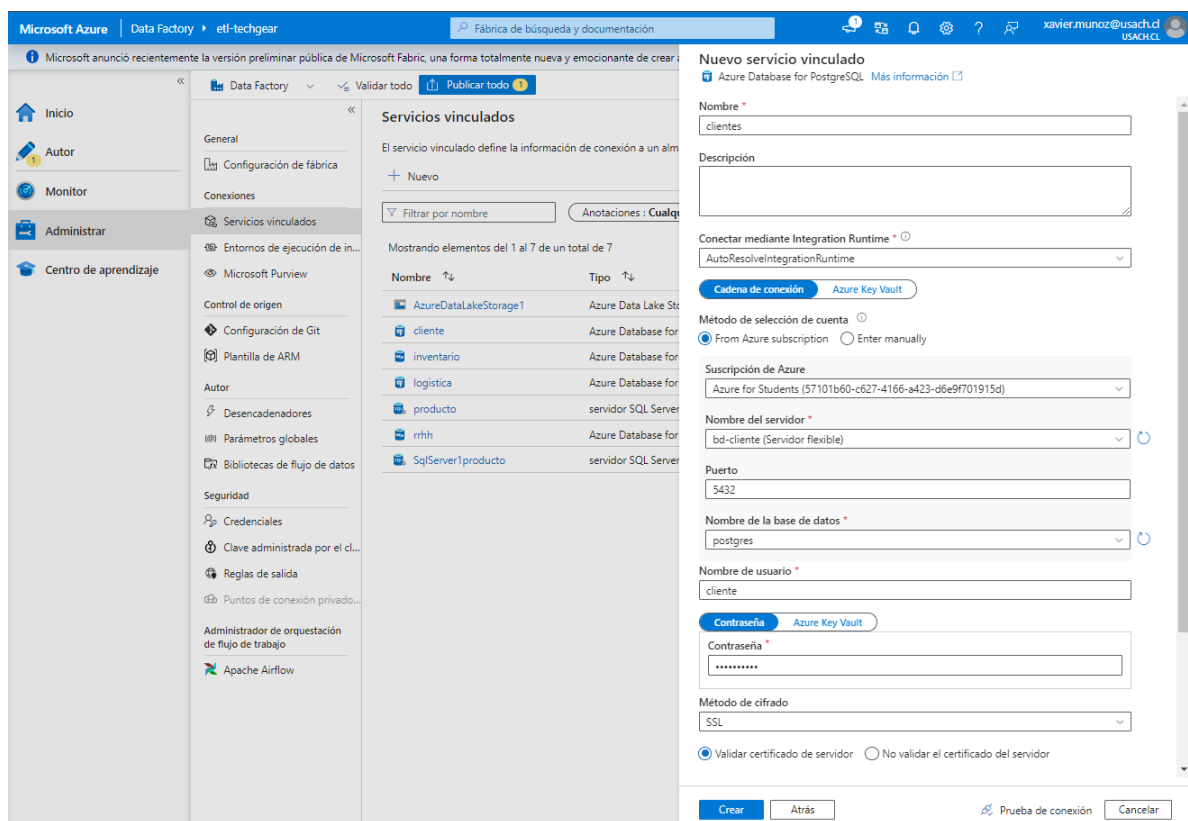


Figura 16: Ejemplo de llenado.

9. Para crear las bases de datos de MySQL (rrhh e inventario) y de SQL Server (producto) se debe guiar por el llenado del paso anterior ya que se procede similarmente, seleccionando “Azure Database for MySQL” y “Azure SQL Database” según corresponda.
10. Luego ir en la barra a “Autor” (símbolo de lápiz), conjunto de datos, nuevo conjunto de datos, buscar y seleccionar “Azure Data Lake Storage Gen2” y dar clic en “Continuar”, luego seleccionar “Delimitedtext” y dar clic en “Continuar”. Ahora en el apartado “Establecer propiedades” se debe ingresar como nombre “csv”+ “nombre del contenedor”, en “Servicio vinculado” seleccionar “AzureDataLakeStorage1”, escribir la ruta donde se guardará el archivo y luego dar clic en “Aceptar”. Un ejemplo de este procedimiento se muestra en la Figura 17. En nuestro caso, se realiza el paso anterior hasta crear los archivos csvBoleta, csvProducto y csvProveedor.

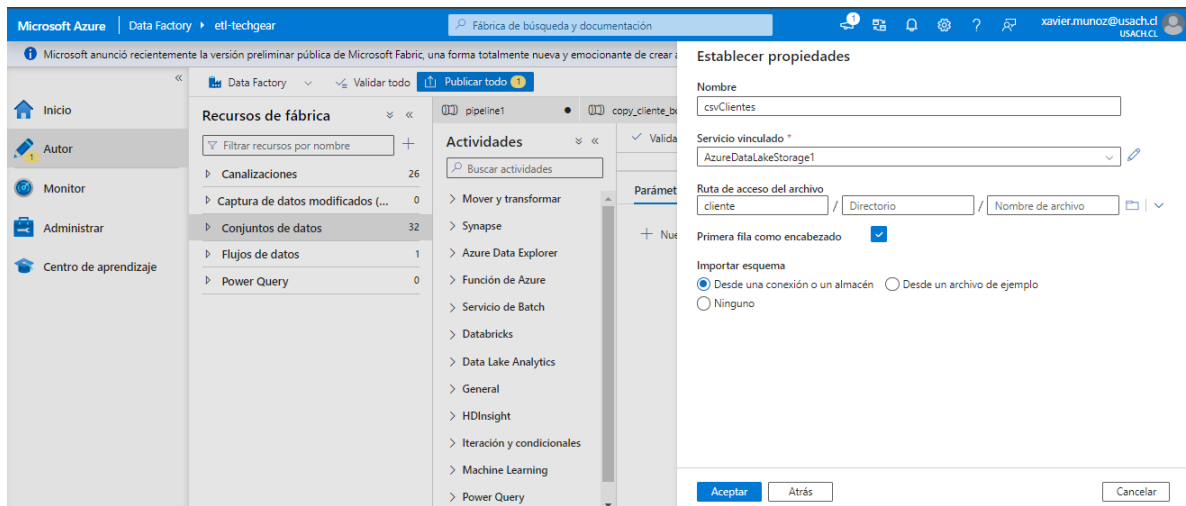


Figura 17: Ejemplo de propiedades.

11. Para crear el archivo JSON, se procede de la misma manera que el paso anterior con la diferencia que en vez de seleccionar “Delimitedtext” se debe elegir “JSON”. En nuestro caso se realiza para crear el archivo “JsonSucursal”.
12. Para crear los dataset de PostgreSQL, ir a “Autor” (símbolo de lápiz), “Conjunto de datos”, seleccionar el icono de 3 puntos y hacer clic en “Nuevo conjunto de datos”, buscar y seleccionar “Azure Database for PostgreSQL”, en “Establecer propiedades” en la sección “Nombre” ingresar “nombre de la base de datos”+”_”+”nombre de la tabla”, en la sección “Servicio vinculado” seleccionar el nombre de la BD, en Nombre de la tabla” seleccionar el nombre de la tabla y luego aceptar para crear el conjunto de datos. Este procedimiento se debe repetir para todas las tablas de la BD de PostgreSQL.

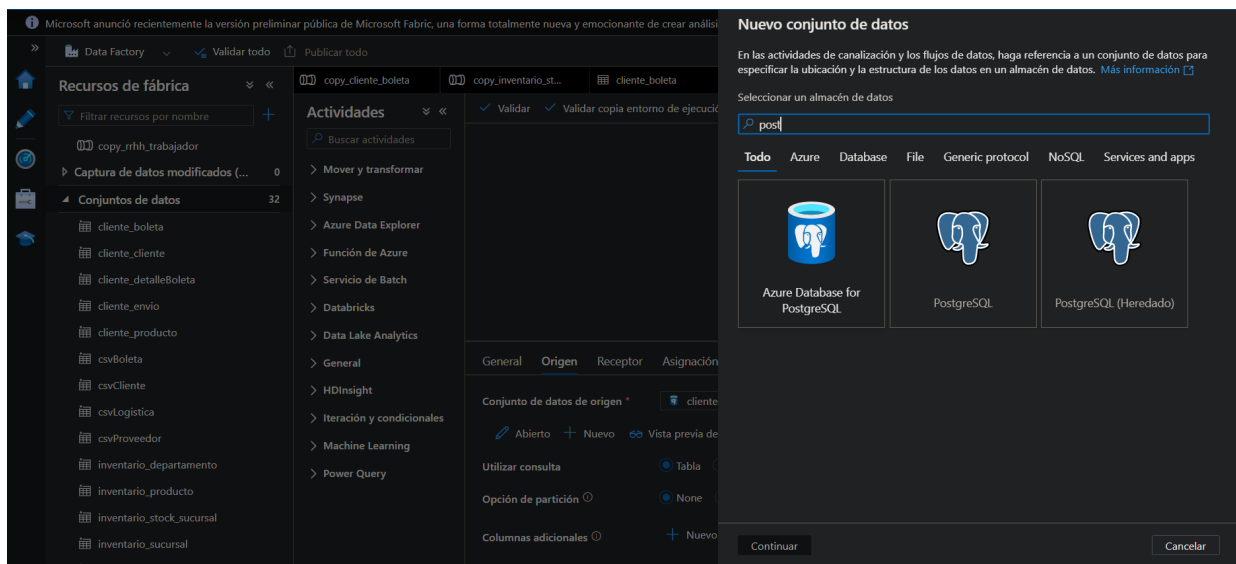


Figura 18: Creación de nuevo conjunto de datos.

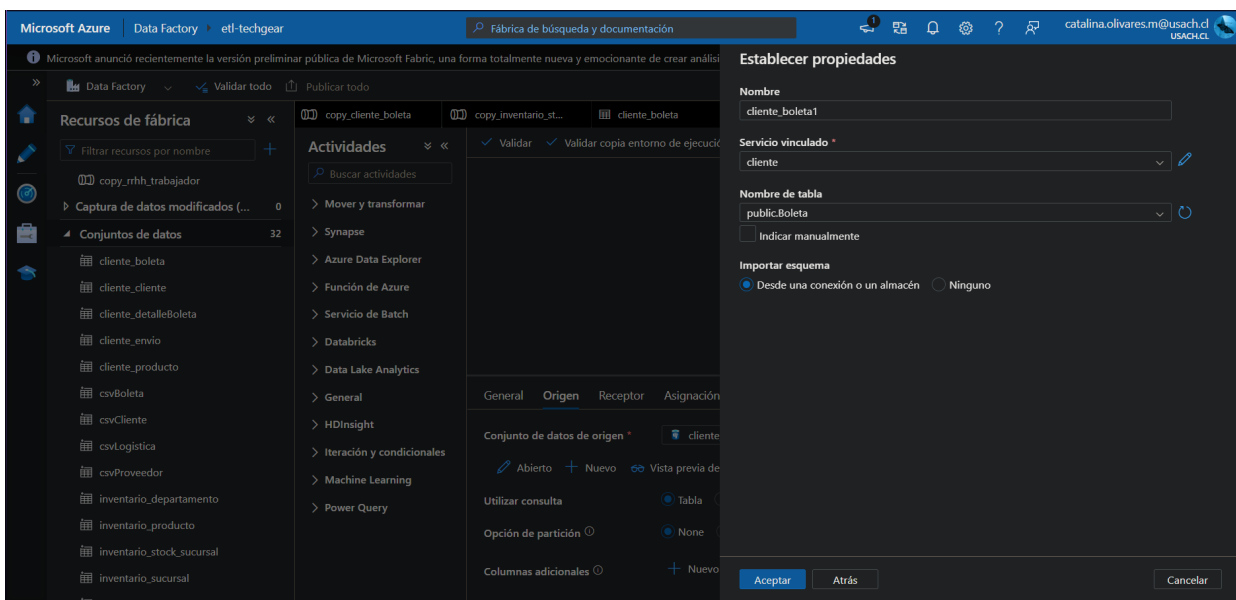


Figura 19: Establecer propiedades

13. Para los datasets de MySQLy SQL Server se debe realizar el paso anterior de la misma forma cambiando el tipo de base de datos por “Azure Database for MySQL” y “Azure SQL Database” según corresponda.
14. Para crear los pipelines, se debe ir al apartado a “Autor” (simbolo de lápiz), seleccionar “Canalizaciones”, dar click a los 3 puntos y seleccionar “Nueva canalización” darle

un nombre al pipeline con el formato “Copy_nombreBD_tabla”, Seleccionar “Mover y transformar” y arrastrar “copiar datos” a la pagina, dar click en “General” y en el apartado “Nombre” ingresar “Copy “+” nombre de la base de datos “+” nombre de la tabla”. Luego ir al apartado “Origen” y seleccionar el dataset “nombre de la base de datos ”+ “nombre de la tabla”, en “Receptor” ingresar “csvBoleta”, ir a “Extensión del archivo” y cambiar por “.csv” y finalmente dar click en depurar.

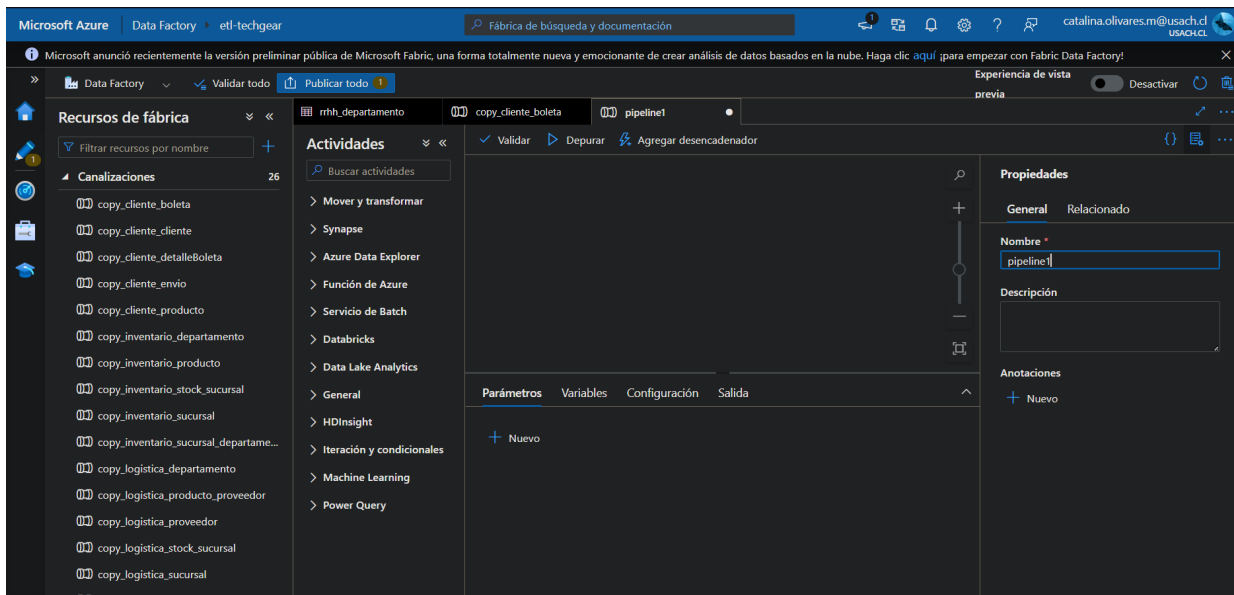


Figura 20: Creación de pipeline.

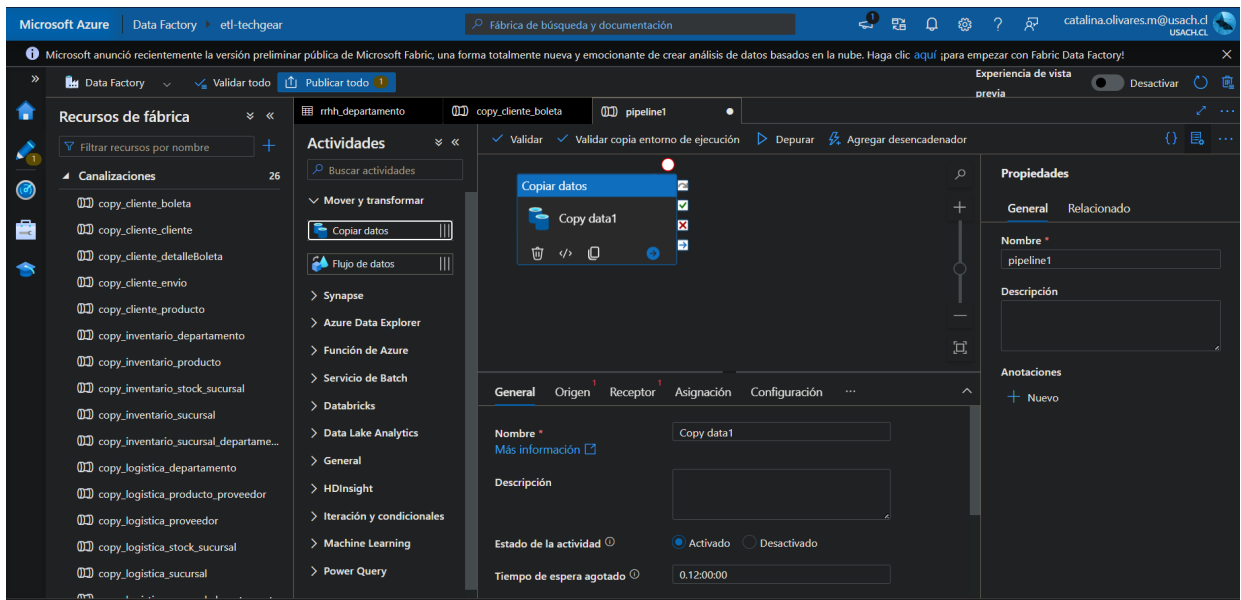


Figura 21: Configuración del pipeline.

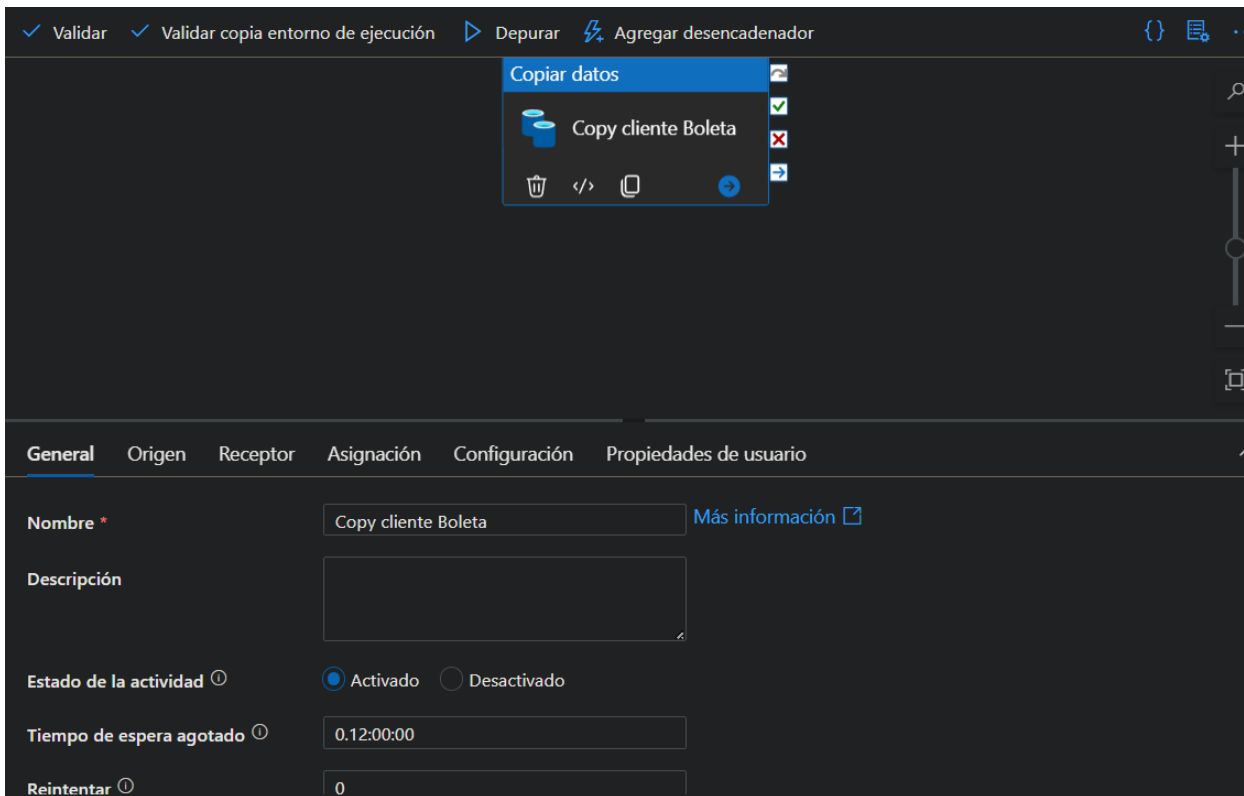


Figura 22: Configuración del pipeline 2.

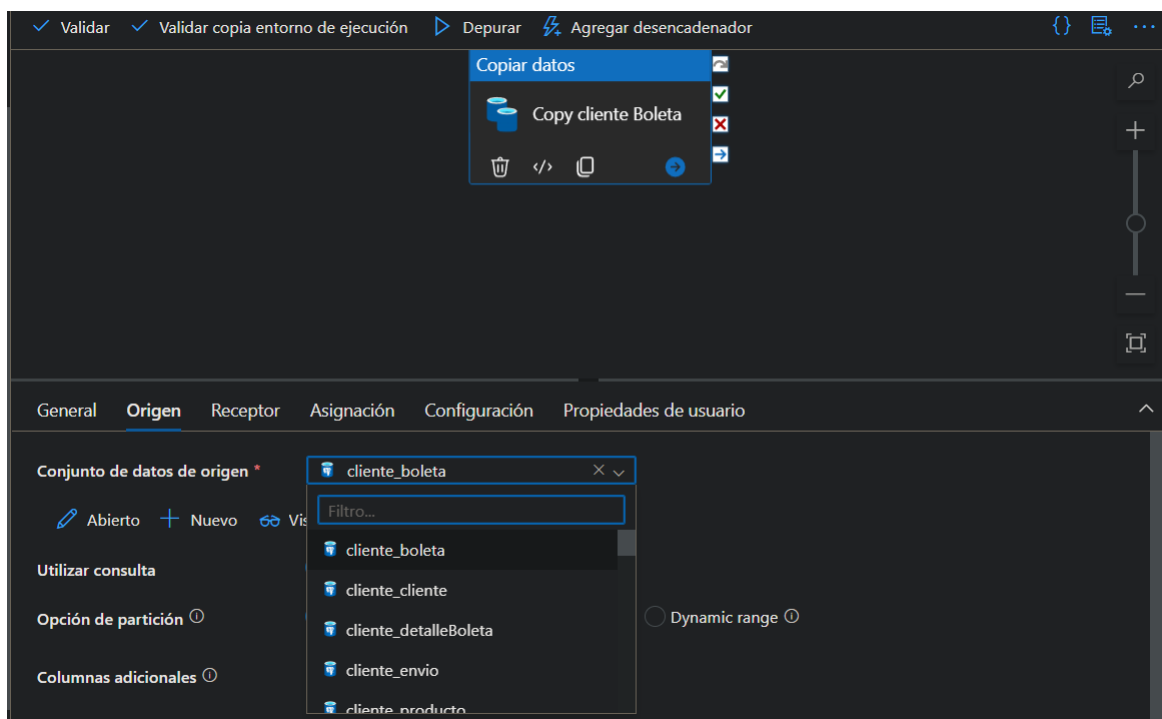


Figura 23: Configuración del pipeline 3.

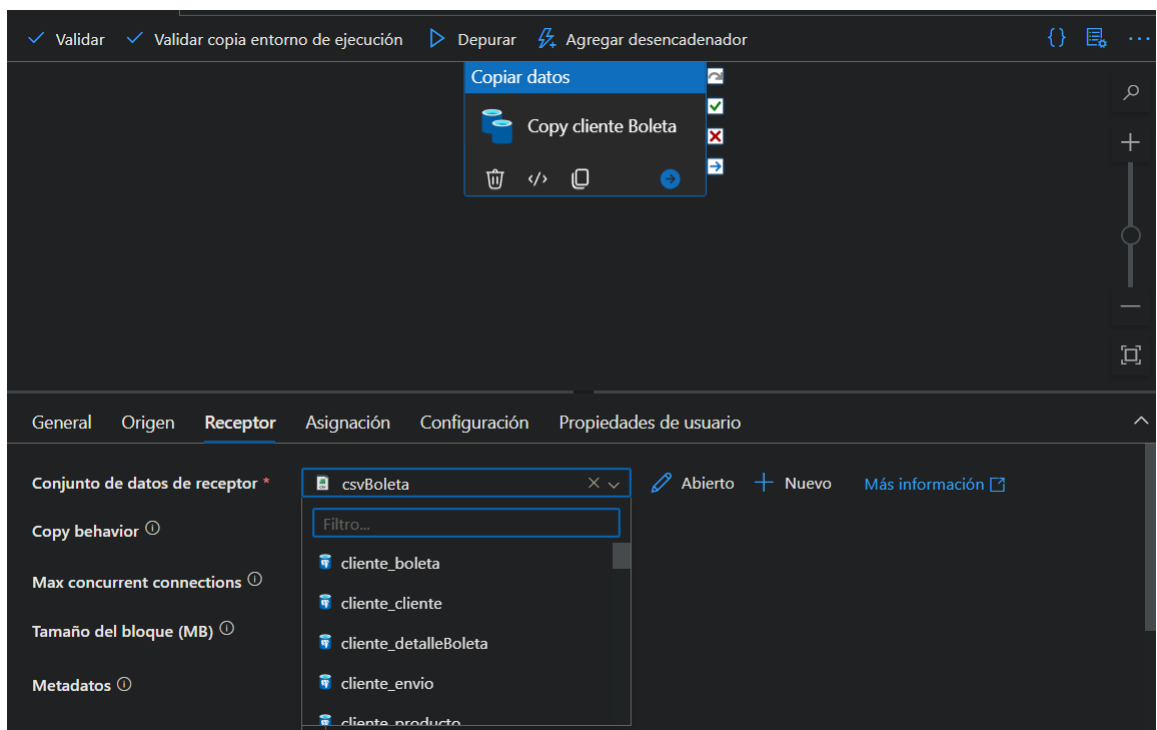


Figura 24: Configuración del pipeline 4.

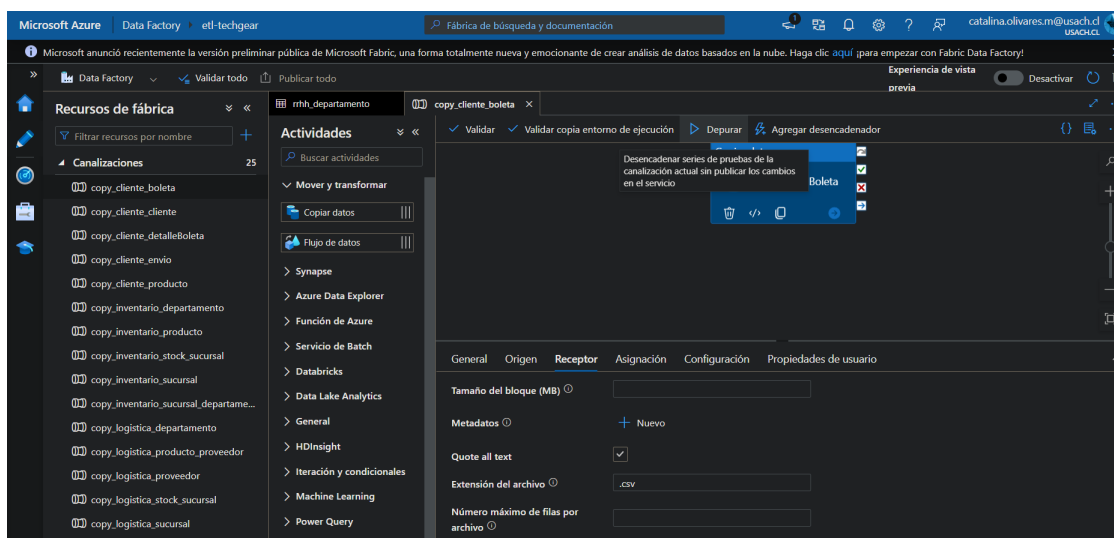


Figura 25: Configuración del pipeline 5.

15. En nuestro caso, en la creación de los pipelines se repite el paso anterior considerando que para los dataset “producto_producto”, “logistica_producto” e “inventario_producto” se selecciona como receptor “csvProducto” y para los dataset “producto_proveedor”, “logistica_producto_proveedor” y “logistica_proveedor” se selecciona como receptor “csvProveedor”. Por otro lado, para los pipelines que tienen como receptor el archivo “JsonSucursal” se debe realizar el paso anterior pero sin ingresar la extensión del archivo. Los datasets utilizados para crear los pipelines con este tipo de archivo son: “producto_sucursal”, “producto_stock_sucursal”, “logistica_dpto”, “logistica_stock_sucursal”, “logistica_sucursal”, “logistica_sucursal_dpto”, “inventario_dpto”, “inventario_stock_sucursal”, “inventario_sucursal”, “inventario_sucursal_dpto”, “rrhh_dpto”, “rrhh_sucursal”, “rrhh_sucursal_dpto” y “rrhh_trabajador”.

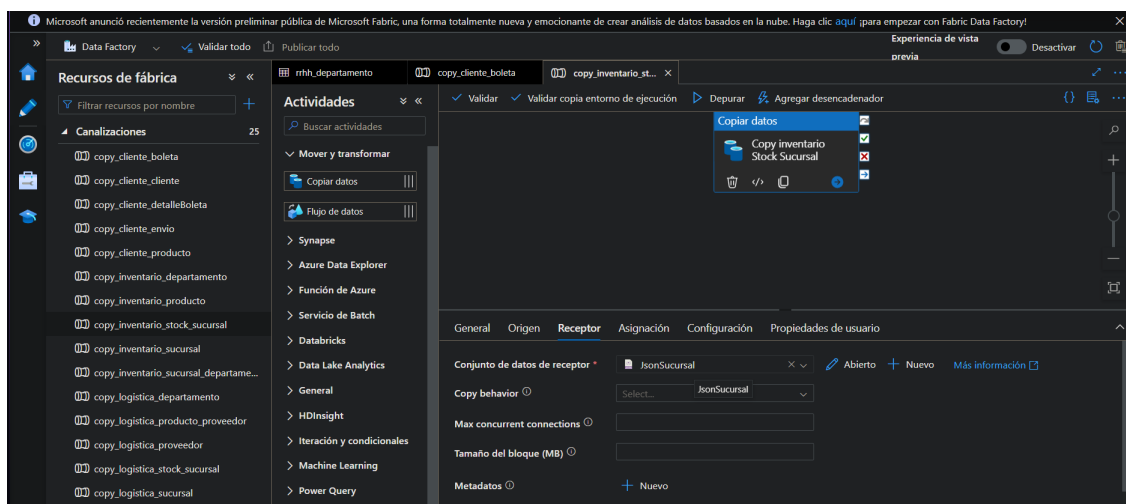


Figura 26: Pipeline con copy_inventario_stock_sucursal.

3. Power BI

1. Seleccionar obtener datos desde Azure Blob Storage.

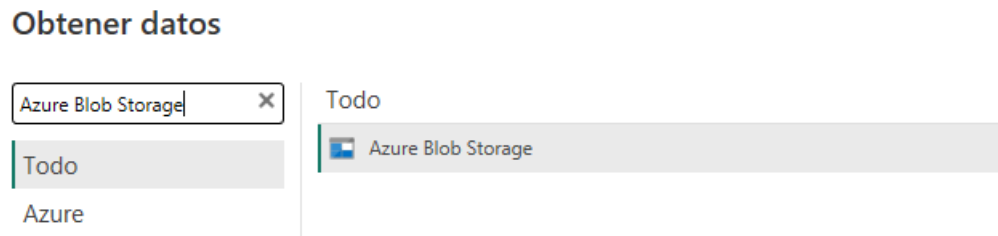


Figura 27: Opción de obtener datos en Azure Blob Storage.

2. Para realizar la conexión, se utilizan las credenciales:

Nombre de la cuenta: `storageaccountteachgear`

Clave 1: `Hn9H5XgV4ssA47Aovu7BKLtoAro0QuUtKI7/oFd06IbTynJdHskj2rk
HT7WnYcI1uTMoeuIWljWC+AStLYy6Ng==`

Clave 2: `dQrXTIdAN73GKuDBktN866unQ7uF`

3. Al realizar la conexión, se pueden seleccionar los contenedores para la importación de datos.

Navegador

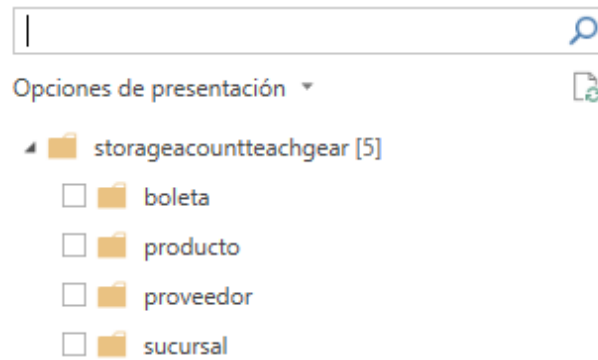


Figura 28: Selección de contenedores.

4. Una vez que se obtienen los datos, es posible realizar transformaciones y analíticas. En el archivo “.pbix” se entrega un análisis tras las transformaciones correspondientes.

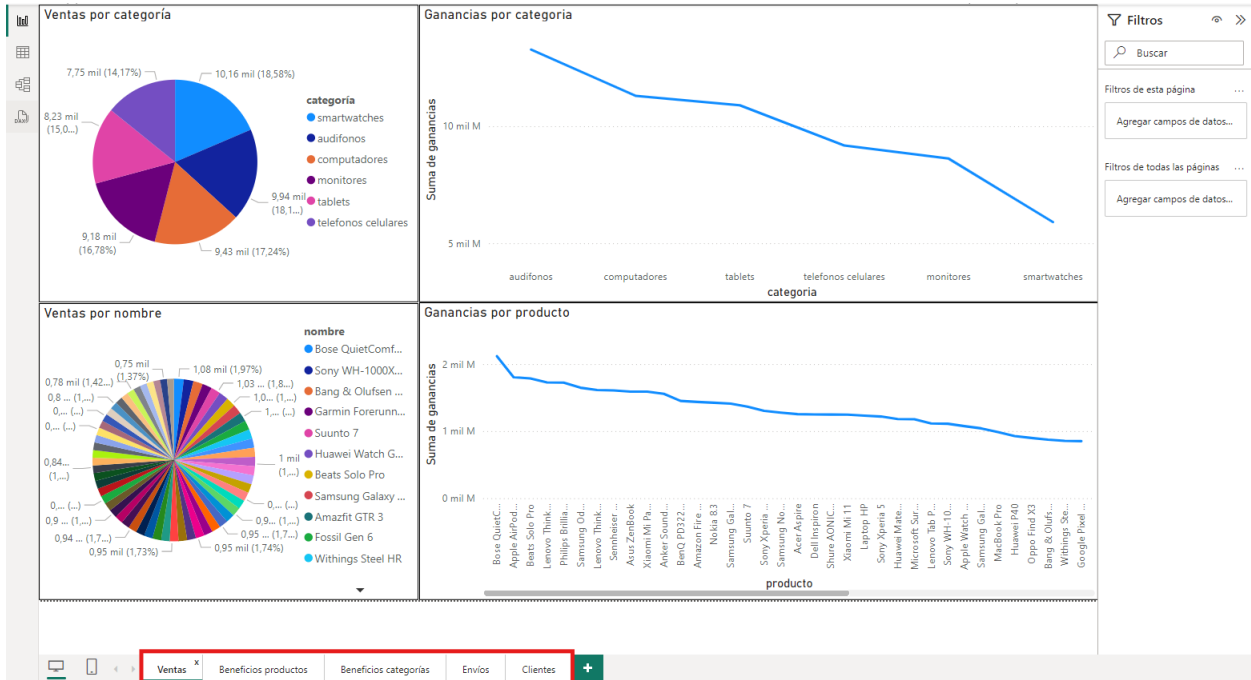


Figura 29: Análisis y gráficos de datos.

5. Se disponen de cinco hojas para navegar en los diferentes análisis de los datos, como se puede apreciar en la Figura 29.