### Creación de un Data Lake en Azure

El objetivo de esta práctica es realizar la ingesta de datos en un **Data Lake** en Azure. Se trasladará el archivo clientes.csv desde un contenedor en **Blob Storage** (ventasbs) hacia un contenedor en **Data Lake** (dataxlake), siguiendo una estructura de zonas.

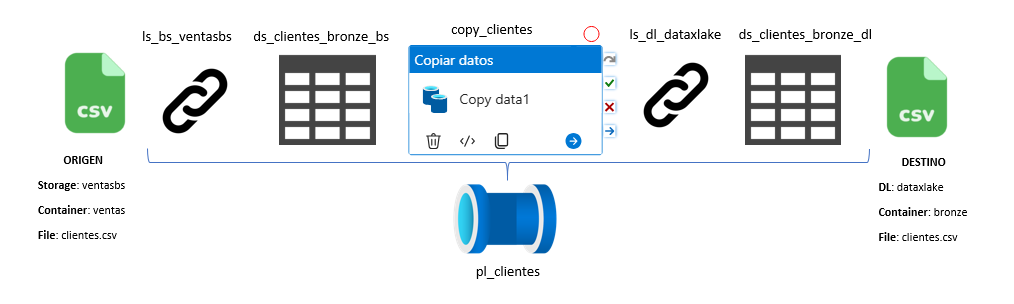


Ilustración 1. Arquitectura del Data Lake en Azure

* **Configuración del Entorno.** Para el almacenamiento de datos en bruto se crearon los siguientes recursos en Azure:
  + Blob Storage

Storage Account: ventasbs

Container: ventas

**Archivo:** clientes.csv

* + Data Lake Storage

Storage Account: dataxlake

Container: bronze

**Archivo:** clientes.csv

* **Preparación de los Contenedores de Almacenamiento.** Se crearon los contenedores requeridos para organizar los datos de manera estructurada:
  + Contenedor ventas en ventasbs (para la ingesta inicial).

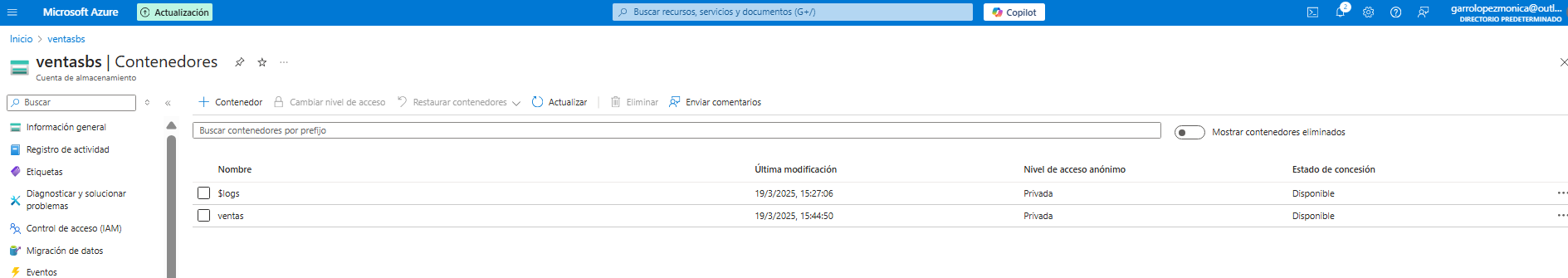


Ilustración 2. Creación del Contenedor ventas en Blob Storage

* + Contenedor bronze en dataxlake (para la fase de almacenamiento en Data Lake).

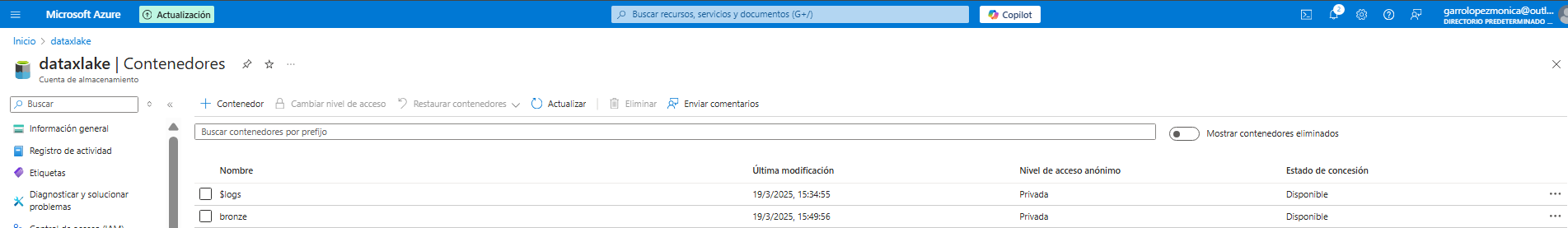


Ilustración 3. Creación del Contenedor bronze en Data Lake

* **Carga de Datos.** El archivo clientes.csv fue cargado en el contenedor de Blob Storage ventas.
* **Configuración de Azure Data Factory.** Para automatizar la transferencia de datos, se utilizó **Azure Data Factory**, y mediante los **Servicios Vinculados (Linked Services),** se crearon conexiones entre los servicios de almacenamiento de Azure.

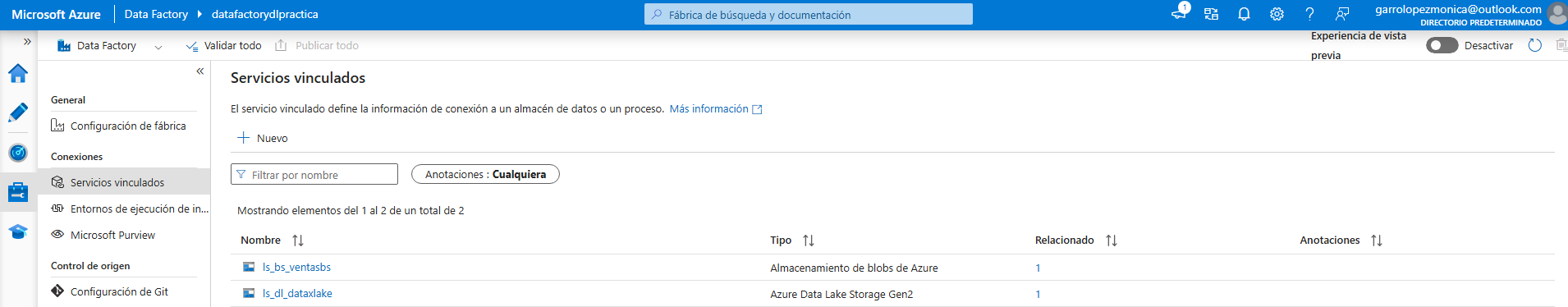


Ilustración 4. Configuración de Linked Services en Azure Data Factory

1. **Conjuntos de Datos (Datasets)**: Se definieron los formatos de datos de entrada y salida y mediante la creación de un **Pipeline** se diseñó un flujo de trabajo para mover los datos desde **Blob Storage** a **Data Lake**.

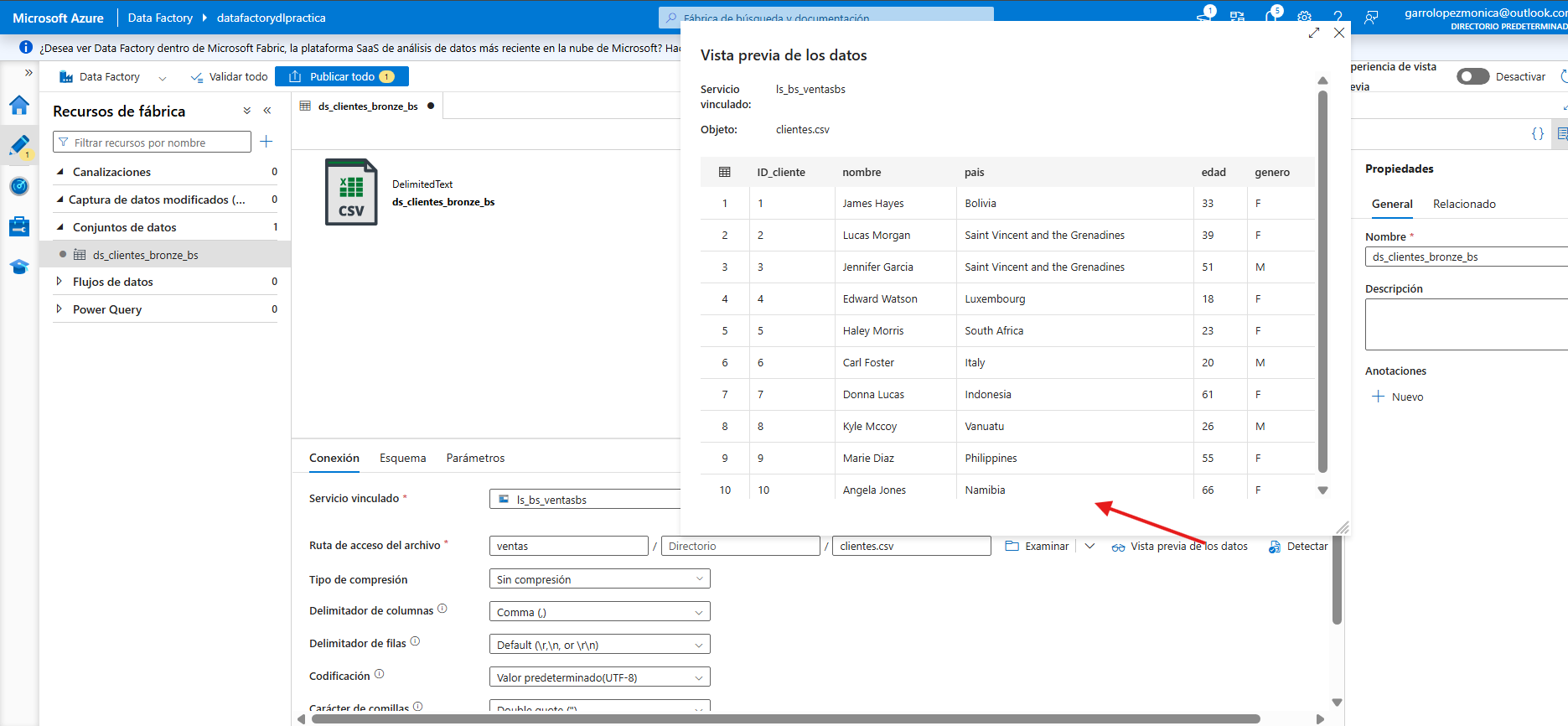


Ilustración 5. Creación del Dataset ds\_clientes\_bronze\_bs en Azure Data Factory

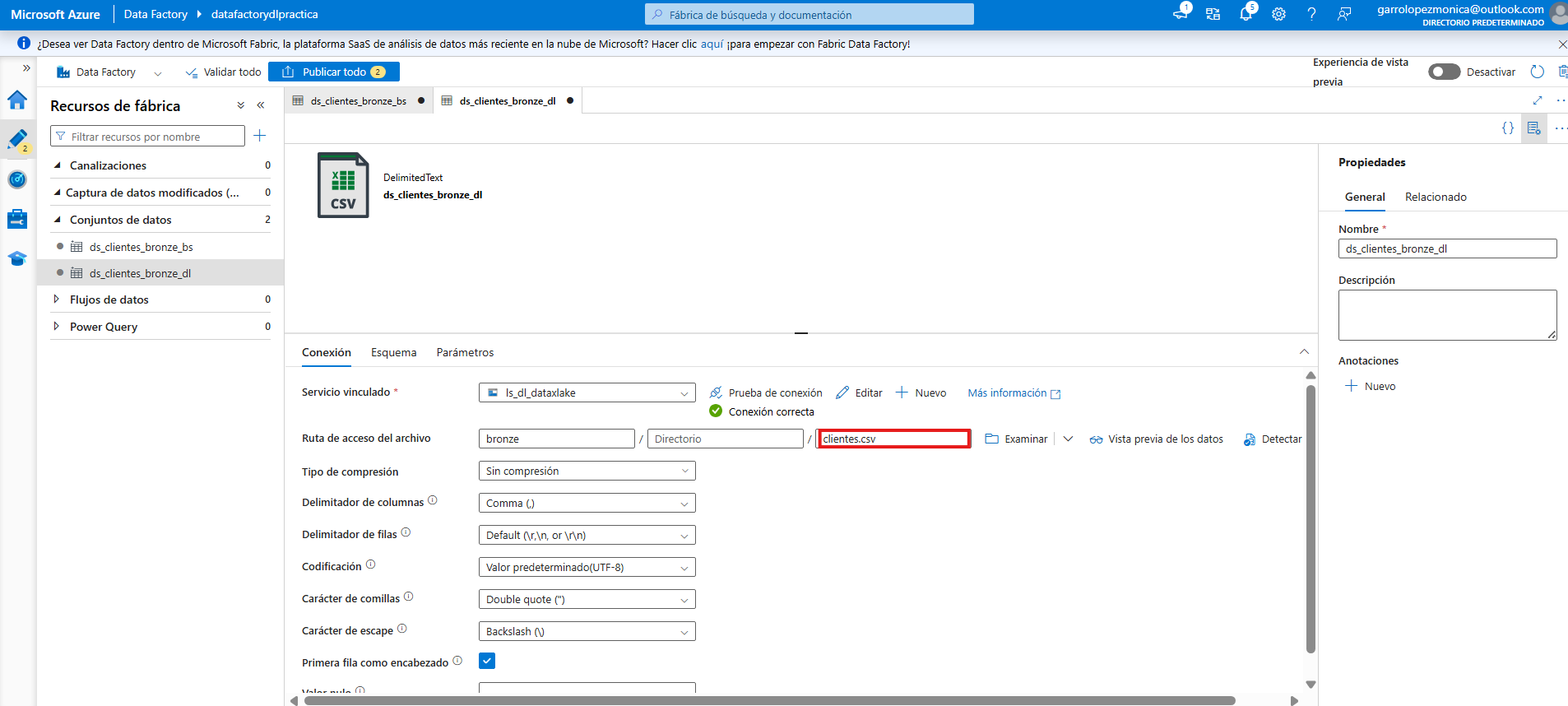


Ilustración 6. Creación del Dataset ds\_clientes\_bronze\_dl en Azure Data Factory

**Ejecución del Pipeline**

Una vez configurado el **pipeline**, se ejecutó el proceso de ingesta, trasladando clientes.csv desde el **Blob Storage** hacia la capa **Bronce** del **Data Lake**.

* Se verificó la correcta ejecución en el **monitor de Data Factory**.
* Se validó la presencia del archivo en dataxlake/bronze.

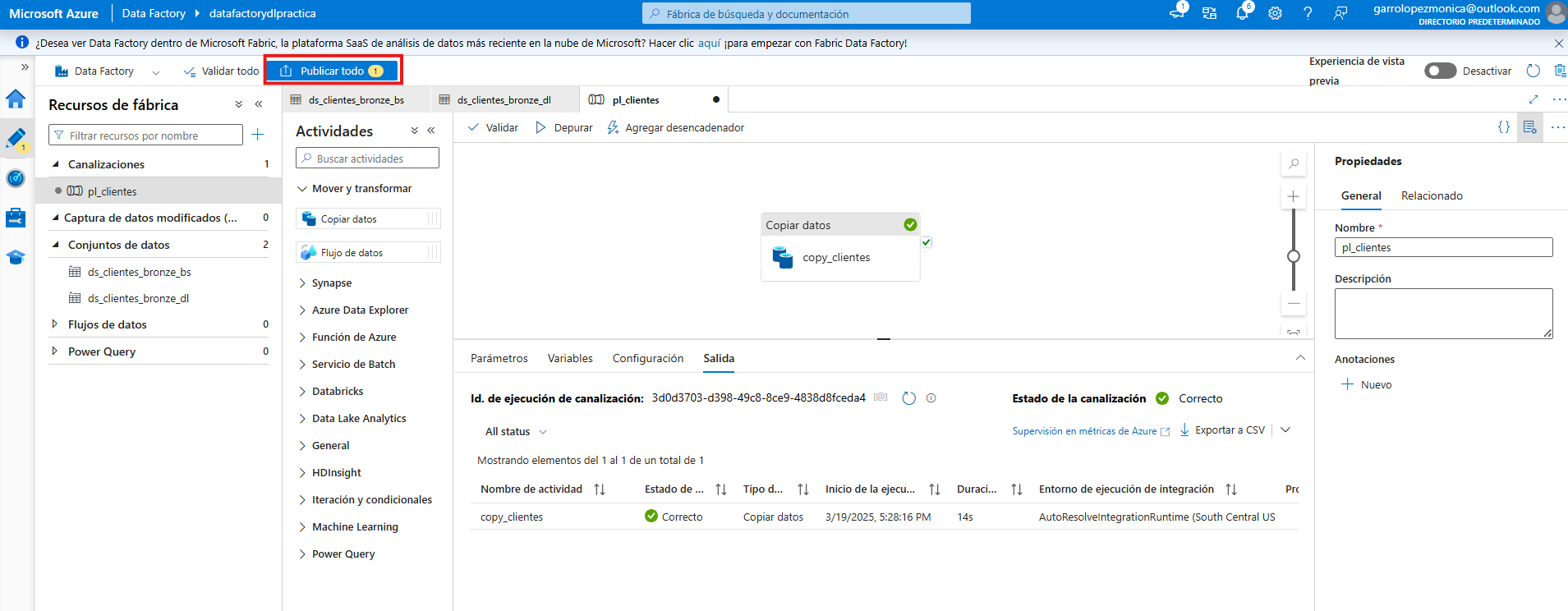


Ilustración 7. Desarrollo del Pipeline de Ingesta de Datos (pl\_clientes)

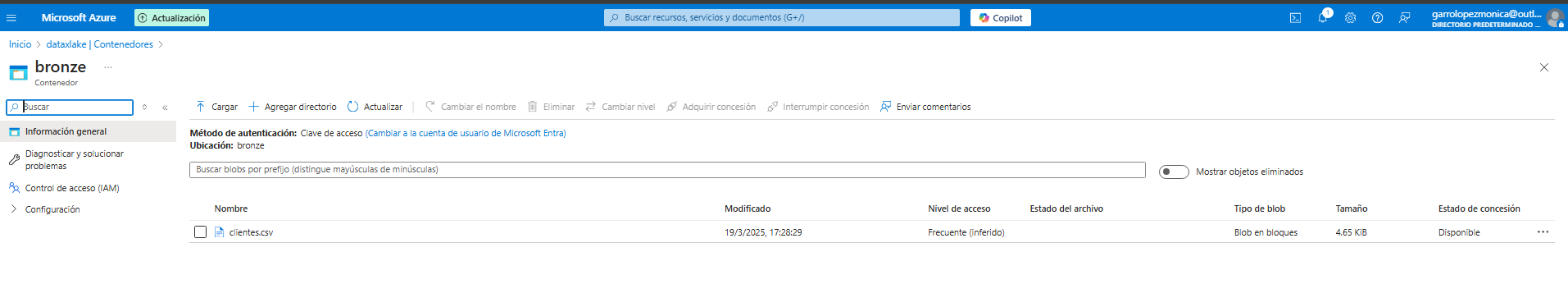


Ilustración 8. Verificación del traslado exitoso del archivo a la capa Bronze