

Laboratorio III: Transformación y Análisis con Synapse Analytics

Introducción

¡Transforma datos crudos en información valiosa! En este laboratorio, utilizarás Azure Synapse Analytics para procesar datos desde la capa Bronze hasta Gold con Spark Pools. Crearás una base de datos SQL serverless y ejecutarás scripts para analizar datos como **covid_worldwide_rc.csv**. Perfecto para aprender procesos ELT y preparar datos para visualización. ¡Descubre el poder de los datos listos para análisis!

Paso 1: Ingresamos a Azure Synapse Analytics

1. Primeramente, ingresamos a la plataforma de Microsoft Azure

Buscar recursos, servicios y documentos (G+/)

Copilot

Servicios de Azure

Crear un recurso

Azure Synapse Analytics

Factorías de datos

Suscripciones

Cuentas de almacenamiento

Grupos de recursos

Máquinas virtuales

Azure Lab Services

Centro de inicio rápido

Más servicios

Recursos

Reciente

Favorito

Nombre	Tipo	Última consulta
syntokyoolympicsgps	Synapse workspace	hace 37 minutos
Azure for Students	Suscripción	hace 15 horas
adftokyoolympicsgps	Factoría de datos (V2)	hace 17 horas
tokio-olympics-gps	Grupo de recursos	hace 17 horas
tokioolympicsgps	Cuenta de almacenamiento	hace 18 horas

Ver todo

2. Ahora ingresamos a la instancia de **Azure Synapse Analytics**.

Microsoft Azure

Buscar recursos, servicios y documentos (G+/)

Copilot

syntokyoolympicsgps

Synapse workspace

Buscar

Información esencial

Grupo de recursos (tagged) : tokio-olympics-gps

Estado : Succeeded

Ubicación : East US 2

Suscripción (tagged) : Azure for Students

Id. de suscripción : 941c6b7-dfb5-4898-8095-7eeae75a3e43

Red virtual administrada : No

Id. del objeto de identidad... : a2e1f6c-9408-41a7-4938-6f02c7f95c

Dirección URL web del al... : https://web.azure.synapse.net/es/Workspace%3FSubscription%3D941c6b7-dfb5-4898-8095-7eeae75a3e43

Etiquetas (tagged) : Azure for Students

Redes : Mostrar configuración del firewall

Dirección URL de la cuenta... : https://tokioolympicsgps.dfs.core.windows.net

Sistema de archivos de A... : dataazkioolympicsgps

Nombre de usuario del a... : rgadminuser

Administrador de Micros... : golan@bucentraur.ac

Punto de conexión de S... : syntokyoolympicsgps.sql.azure.synapse.net

Punto de conexión de S... : syntokyoolympicsgps-ondemand.sql.azure.synapse.net

Punto de conexión de de... : http://syntokyoolympicsgps.dev.azure.synapse.net

Abrir Synapse Studio

Empiece a crear su solución de análisis completamente integrada y descubra información nueva.

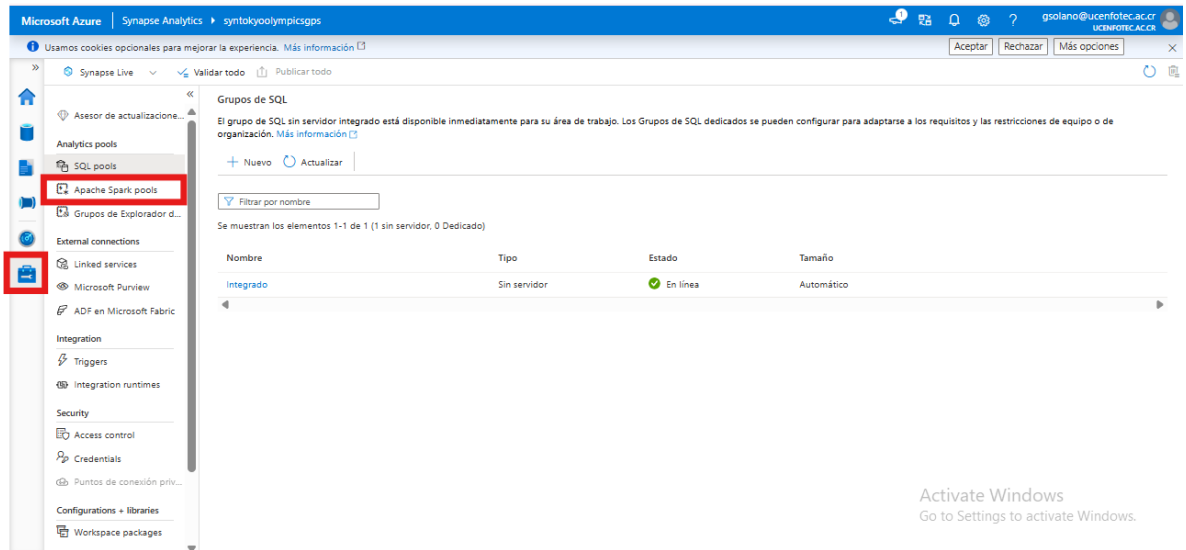
Abrir

Lectura de documentación

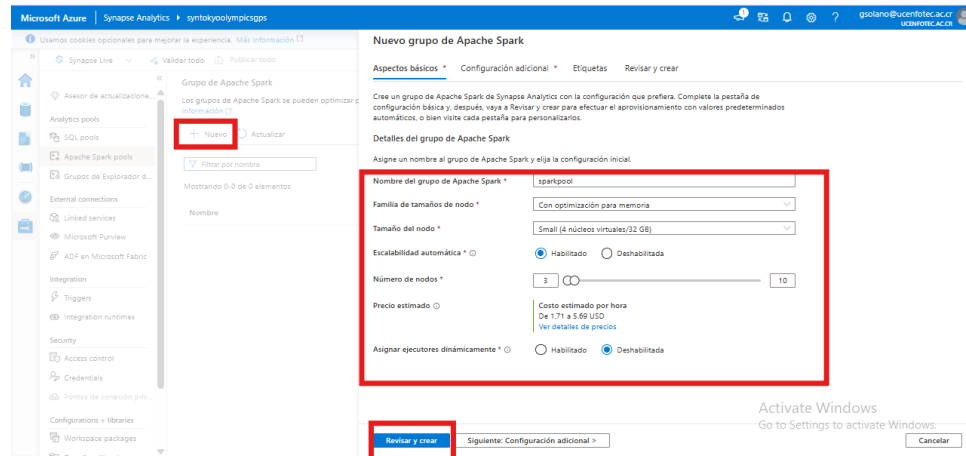
Obtenga información sobre cómo ser productivo rápidamente. Explore los conceptos, los tutoriales y los ejemplos.

Más información

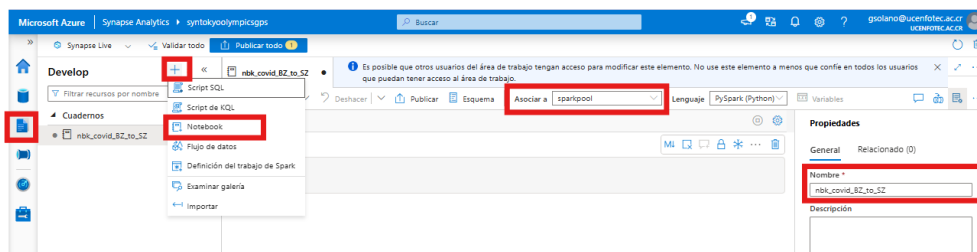
3. Procedemos a ingresar a en la pestaña de **manage** y luego ingresamos a **Apache Spark Pools**.



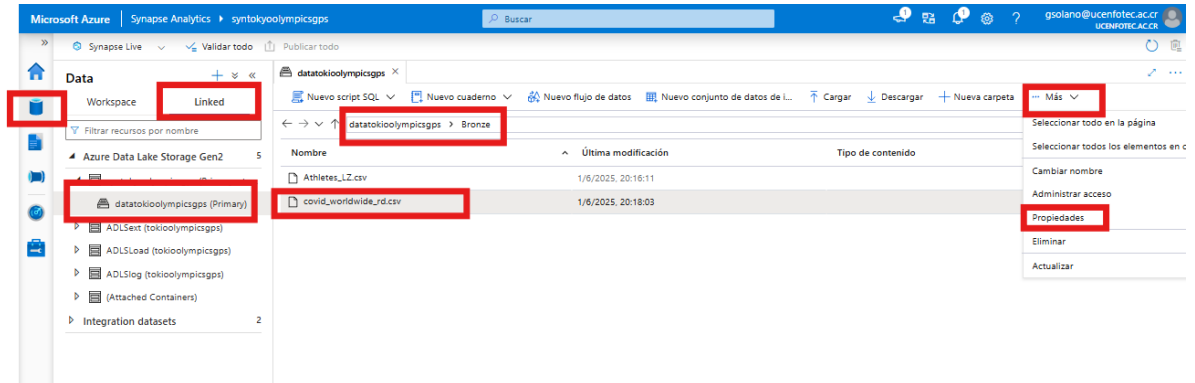
4. Procedemos a crear un nuevo spark pool con el nombre sparkpool, con optimización para memoria que es el único que nos permite crear por la licencia de estudiante, la familia de los nodos seleccionamos small y procedemos a crear el nuevo pool.



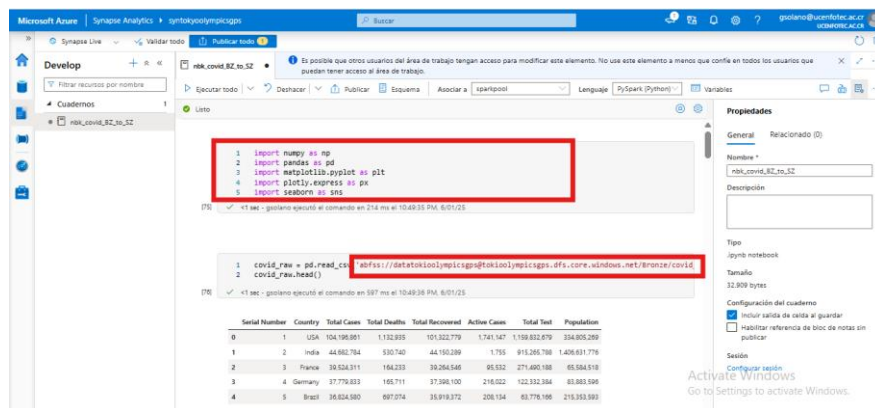
5. Luego ingresamos a **development** y creamos un nuevo notebook al cual nombramos **nbk_covid_BZ_to_SZ** y seleccionamos el sparkpool que creamos anteriormente.



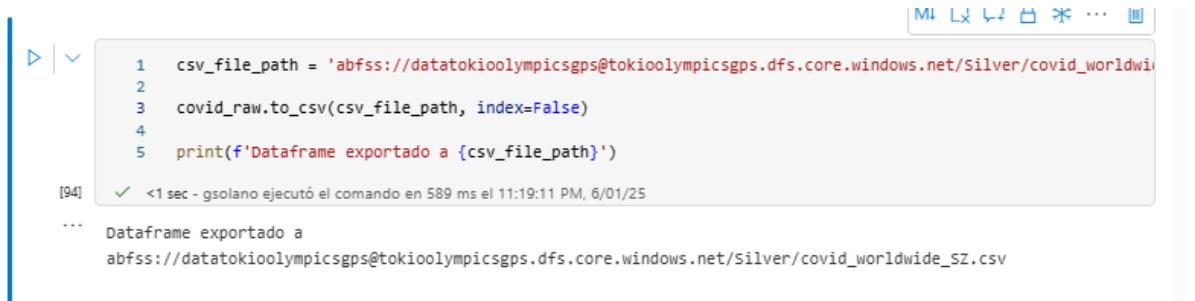
6. Ahora vamos a **data, linked**, seleccionamos nuestro lago de datos en carpeta de file bronze seleccionamos **covid_worldwide_rd.csv** le damos click al botón de **más**, seleccionamos las **propiedades**. Seleccionamos la **Ruta de ABFSS**.



7. Regresamos al notebook, copiamos y pegamos una a una de las chunks del notebook que se encuentra en la **Evaluación** (semana 3, nbk_covid_BZ_to_SZ), Además, cambiamos el **Ruta de ABFSS**, que es única para cada uno.



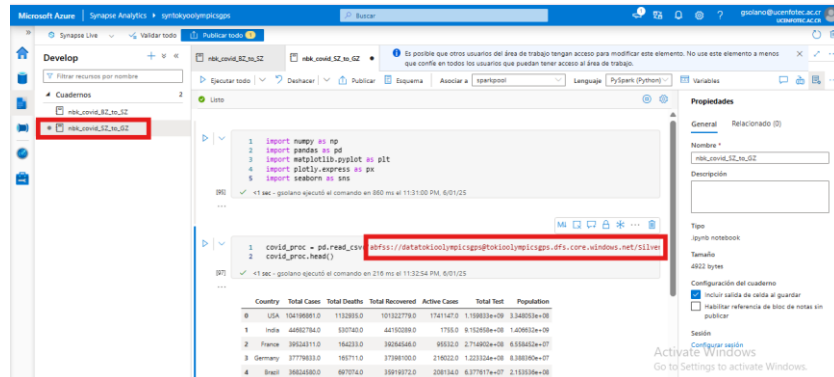
8. Al final del notebook encontramos el envío de nuestros datos al la carpeta Silver, modifique el path de acuerdo con la carpeta que creo en el laboratorio 1.



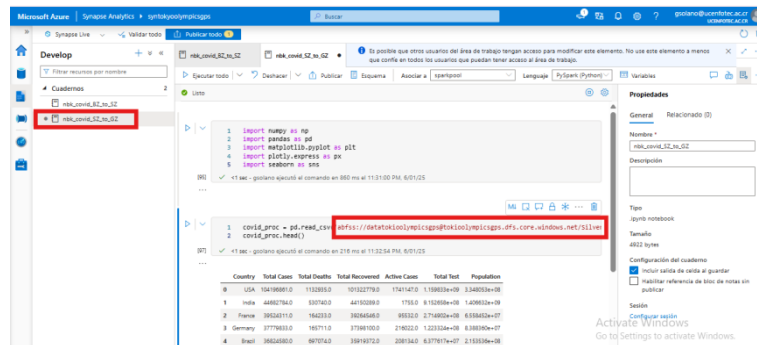
Por último, procedemos a **publicar todo**.

Paso 2: ingresamos a development

1. Creamos un nuevo notebook al cual nombramos **nbk_covid_SZ_to_GZ** y seleccionamos el sparkpool que creamos anteriormente y cambiamos la ruta de donde se encuentra el archivo que realizamos en el **paso 1** que se encuentra en la carpeta **Silver**.

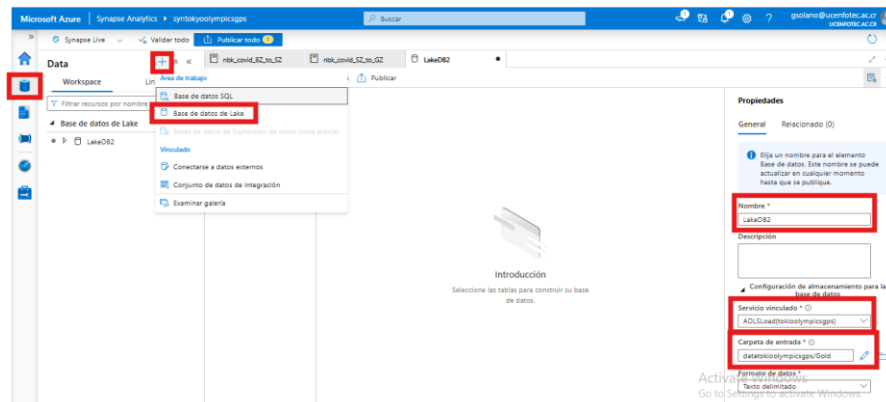


2. Copiamos y pegamos una a una de las chunks del notebook que se encuentra en la **Evaluación** (semana 3, nbk_covid_SZ_to_GZ),

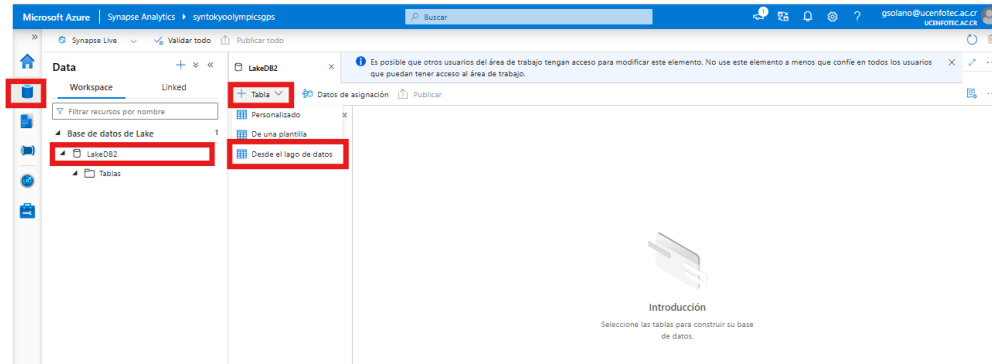


Paso 3: Synapse Data Load

1. Ahora vamos a **data, Workspace** y creamos una **base de datos lake** con el nombre de **LakeDB2**, además seleccionamos el **Servicio vinculado** **ADLSLoad(tokioolympicsgps)** y la **Carpeta de entrada** **datatokenioolympicsgps/Gold**, por último, procedemos a **publicar todo**.



- Luego procedemos a **abrir el LakeDB2**, desde la tabla seleccionamos **Desde el lago de datos**. Le asignamos un nombre **TBL_ best_recovery_GZ**, y **Servicio vinculado (ADLSLoad(tokioolympicsgps))** y **Archivo o carpeta de entrada (datatokioolympicsgps/Gold/best_recovery_GZ.csv)** y procedemos a darle clic a **aceptar**



Creación de una tabla externa a partir del lago de datos

Detalles de la tabla externa

Seleccione la ubicación en donde que se almacenan provisionalmente los archivos que contienen los datos. Se admiten Azure Data Lake Storage (ADLS) Gen2 y Azure Blob Storage. [Más información](#)

Nombre de la tabla externa *	<input type="text" value="TBL_best_recovery_GZ"/>
Servicio vinculado * ⓪	<input type="text" value="ADLSLoad(tokioolympicsgps)"/>
Archivo o carpeta de entrada * ⓪	<input type="text" value="datatokioolympicsgps/Gold/best_recovery_GZ.csv"/>

- Luego seleccionamos en la nueva pestaña que salta: **Primera fila (Inferir nombres de columna)**

Nueva tabla externa

Configuración del formato de archivo de origen
Especifique el formato y el diseño de los datos. [Más información](#)

Ruta de acceso al archivo
datatokioolympicsgps/Gold/best_recovery_GZ.csv

Preview Data

Tipo de archivo
CSV

Terminador de campo ⓪

☐ Editar

Primera fila
☒ Inferir nombres de columna ⓪

Delimitador de cadena ⓪

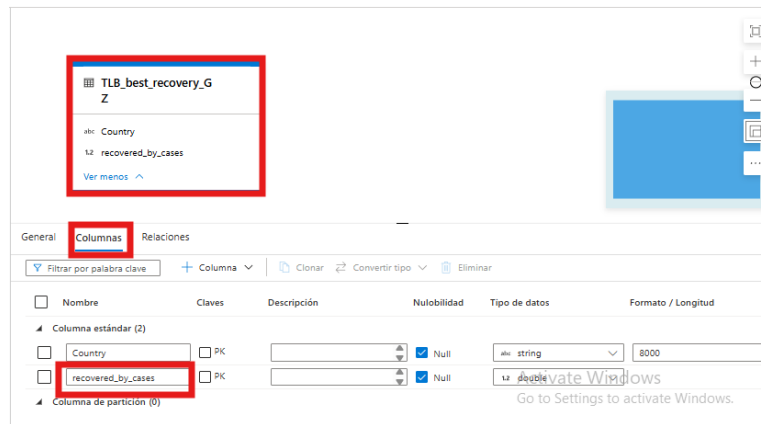
☐ Editar

Use default type ⓪

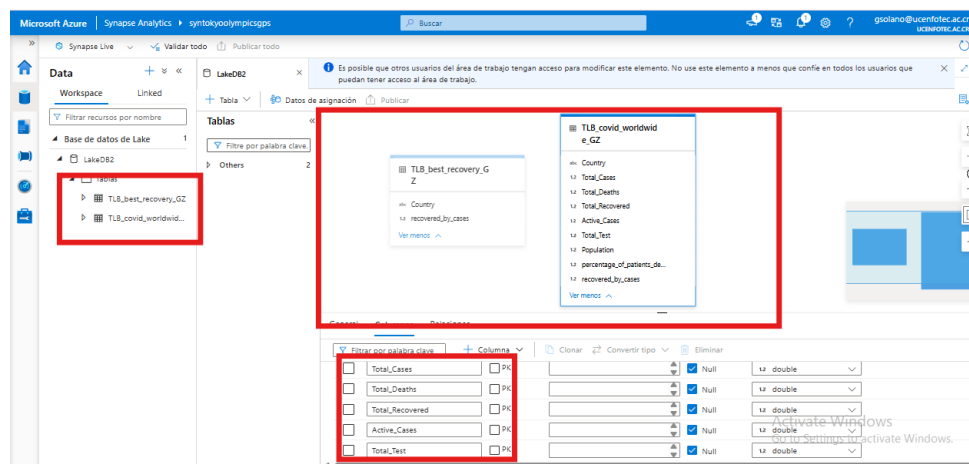
Longitud máxima de la cadena * ⓪

Activate Windows
Go to Settings to activate Windows.

- Por último, seleccionamos la nueva tabla, nos dirigimos a la pestaña de columnas y añadimos guion bajo a los espacios (recovered by cases → recovered_by_cases)

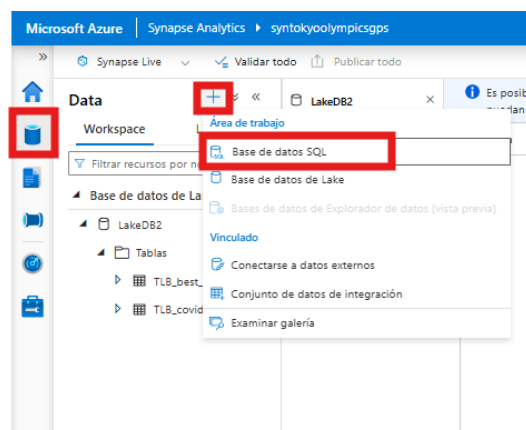


- Repetimos el **paso 3** y creamos la tabla **TLB_best_recovery_GZ** y luego procedemos a **Publicar todo**.

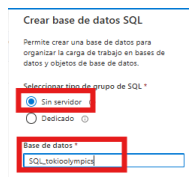


Paso 4: Create SQL Database

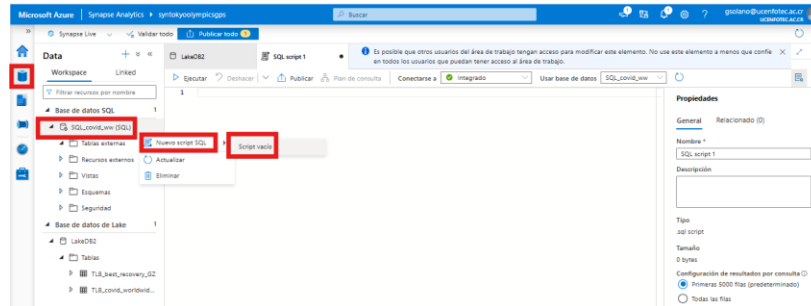
- Creamos la base de datos



El grupo de SQL es **serverless**, con el nombre de **SQL_covid_ww** y procedemos a **crear**.

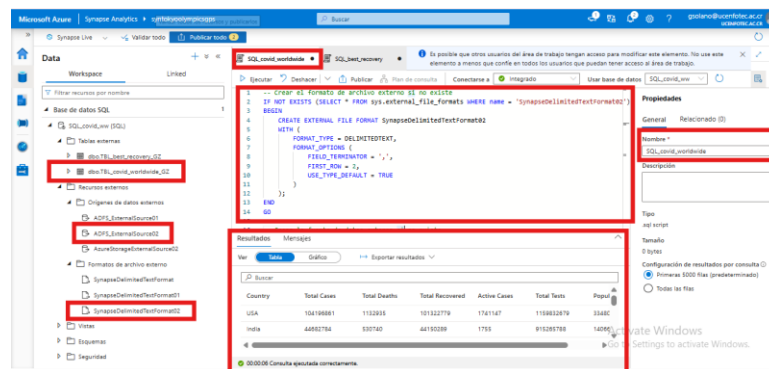


- En la nueva base de datos creado le damos clic a los tres puntos (...) **Nuevo script SQL** y luego clic en **Script vacío**.

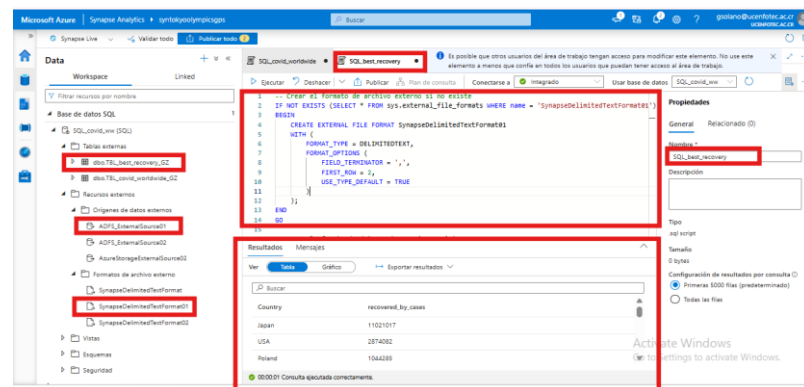


- Vamos a crear dos scripts que se llamen **SQL_covid_worldwide** y **SQL_best_recovery**, los cuales los encontraras los scripts en **evaluación** en la **semana 3**.

3.1 SQL_covid_worldwide



3.2 SQL_best_recovery



- Procedemos a **publicar todo**, fin

YOU DID IT!
CONGRATULATIONS!