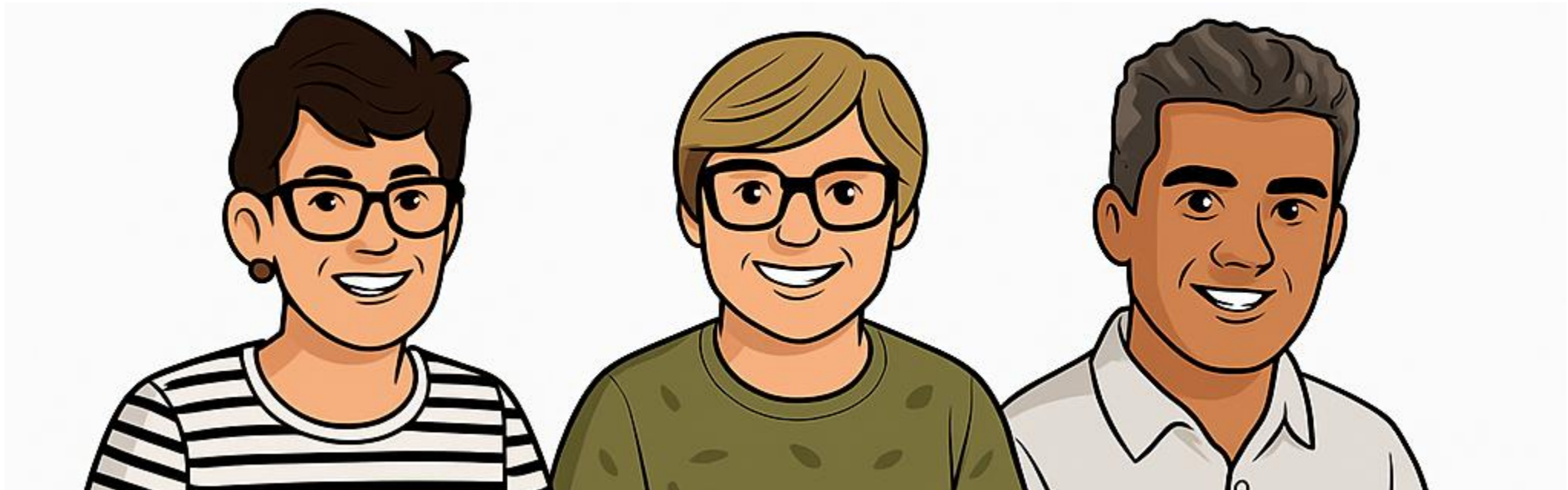


Comenzamos en breve...



A · D · N  **Fabric**

Sesión 12

Transformación de datos en Microsoft Fabric (3)

Ana María Bisbé York
Diana Aguilera Reyna
Nelson López Centeno



¿Quiénes somos?



Consultora y Formadora BI

Autora en :



<https://www.linkedin.com/in/ambynet/>



Consultora y Formadora BI



Fabric Analytics
Engineer Associate

<https://www.linkedin.com/in/diana-aguilera-reyna/>



Consultor y Formador BI



Fabric Analytics
Engineer Associate

<https://www.linkedin.com/in/nelson-lopez-centeno/>



Patrocinadores

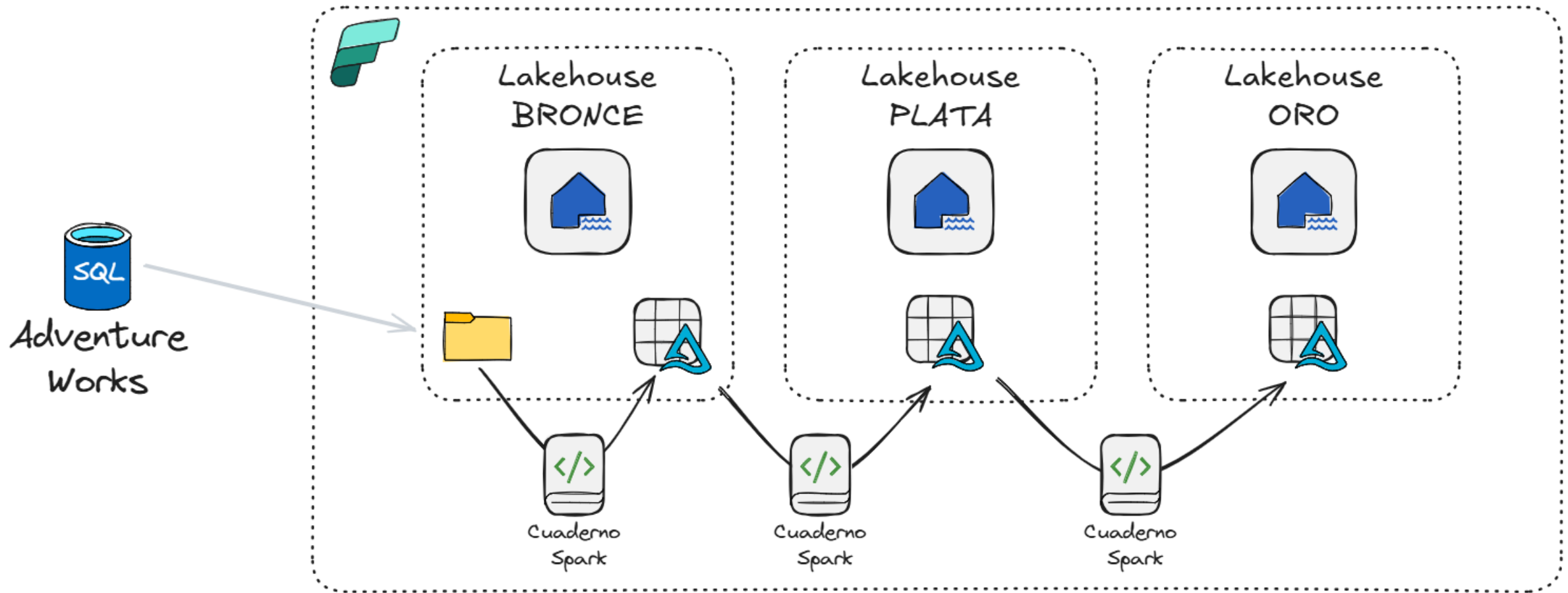


Amby.net

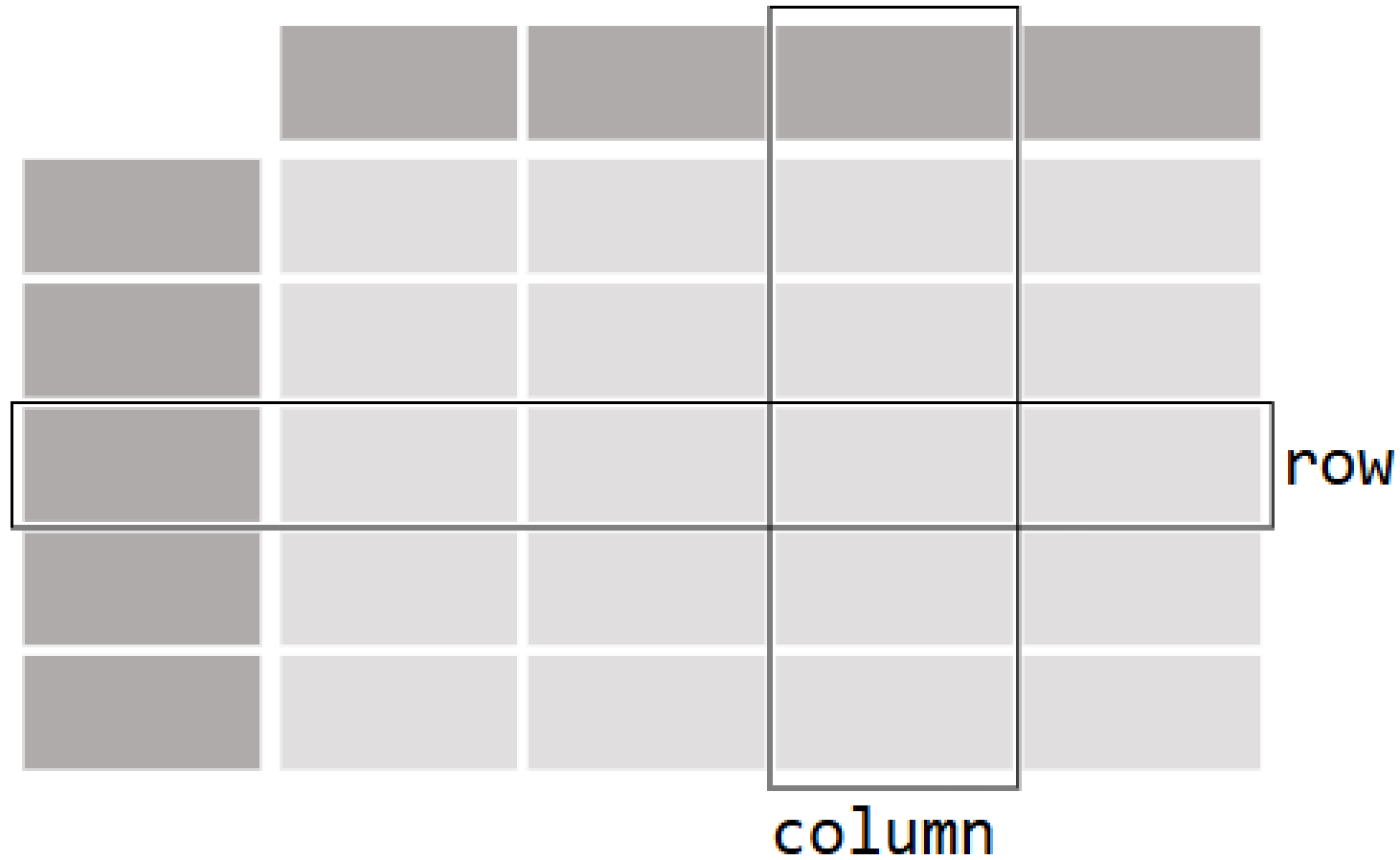
Agenda

- Transformación de datos con PySpark
 - Capa Bronce
 - Capa Plata
 - Capa Oro

Arquitectura Medallón



DataFrame



column

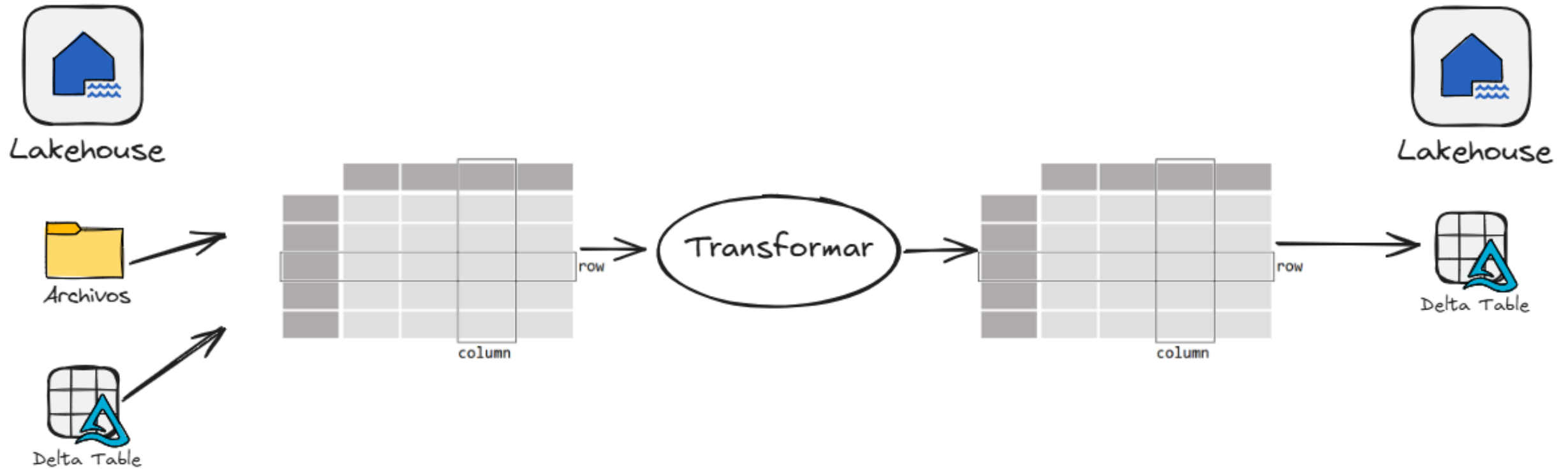
DataFrame



Versión de Spark en Fabric : 3.5



DataFrame



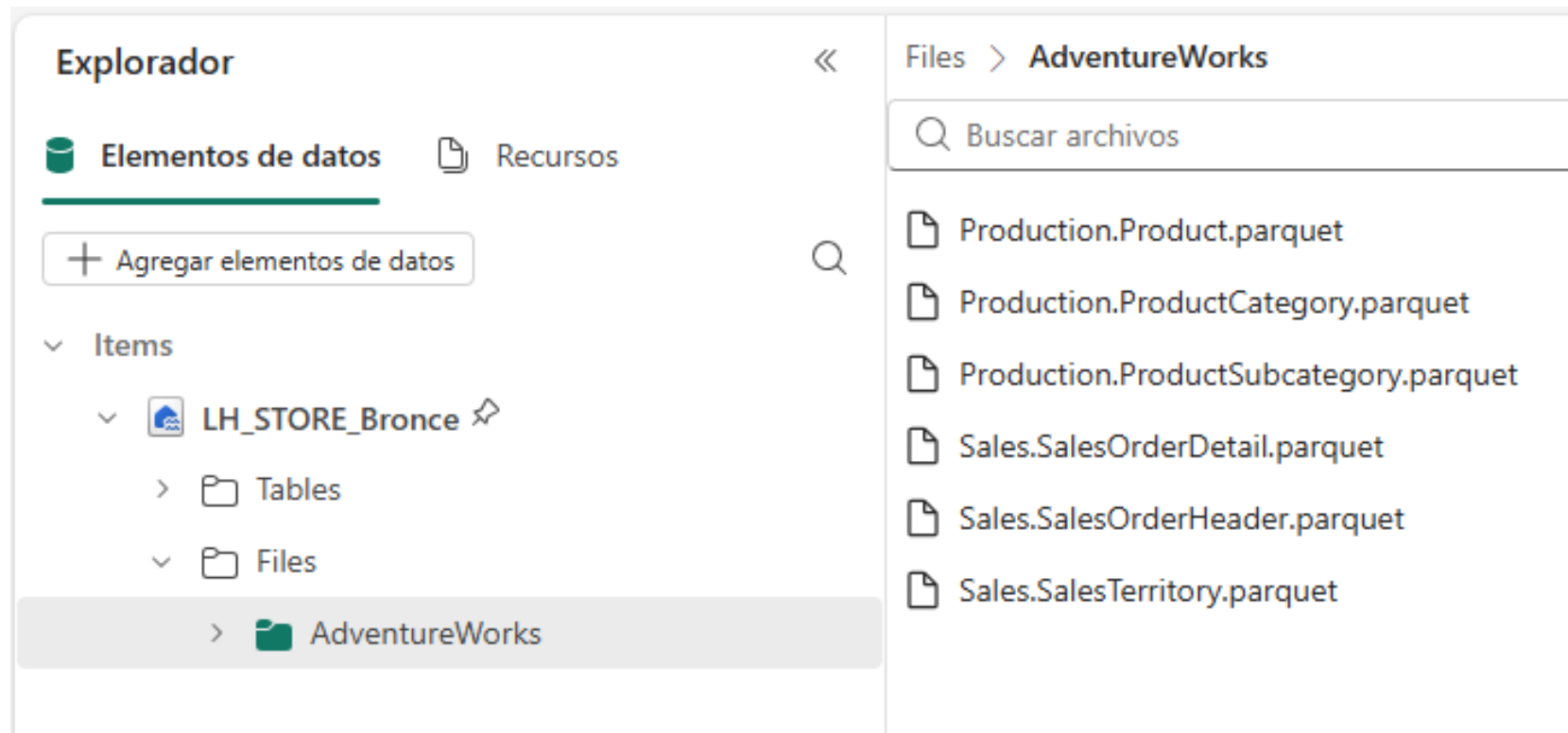
Transformación con ...

Capa Bronce

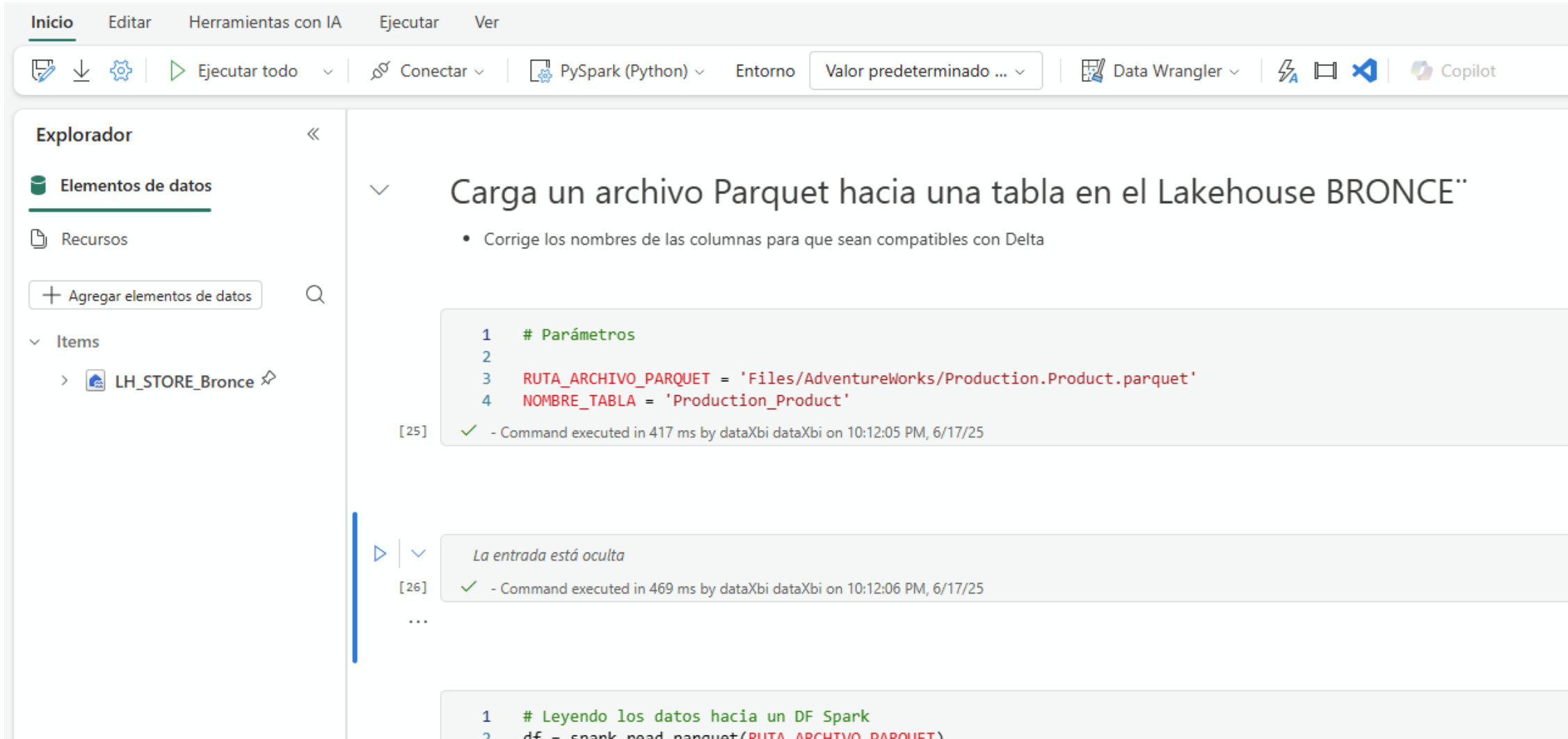
A · D · N  Fabric

Capa Bronce

Archivos Parquet (cargados en la ingesta)



Cuaderno para cargar un archivo Parquet



The screenshot displays the Databricks workspace interface. The top navigation bar includes tabs for 'Inicio', 'Editar', 'Herramientas con IA', 'Ejecutar', and 'Ver'. Below this is a toolbar with icons for file operations, a 'Ejecutar todo' button, a 'Conectar' button, a 'PySpark (Python)' environment selector, an 'Entorno' dropdown, a 'Valor predeterminado ...' dropdown, a 'Data Wrangler' button, and icons for 'Copilot' and other tools.

The left sidebar, titled 'Explorador', shows a tree view of the workspace. It includes 'Elementos de datos', 'Recursos', and a search bar. Under 'Items', there is a folder named 'LH_STORE_Bronce'.

The main area displays a notebook titled 'Carga un archivo Parquet hacia una tabla en el Lakehouse BRONCE'. The notebook content includes a list of instructions: 'Corrige los nombres de las columnas para que sean compatibles con Delta'. Below this, there is a code block with the following Python code:

```
1 # Parámetros
2
3 RUTA_ARCHIVO_PARQUET = 'Files/AdventureWorks/Production.Product.parquet'
4 NOMBRE_TABLA = 'Production_Product'
```

Below the code block, there is a status bar indicating that the command was executed successfully in 417 ms by dataXbi dataXbi on 10:12:05 PM, 6/17/25.

Below the status bar, there is a section titled 'La entrada está oculta' with a status bar indicating that the command was executed successfully in 469 ms by dataXbi dataXbi on 10:12:06 PM, 6/17/25.

Below the status bar, there is a code block with the following Python code:

```
1 # Leyendo los datos hacia un DF Spark
2 df = spark.read.parquet(RUTA_ARCHIVO_PARQUET)
```

Cuaderno para cargar todos los archivos




Inicio


Editar


Herramientas con IA


Ejecutar

Ver




 Ejecutar todo




 Conectar


 PySpark (Python)

Entorno

Valor predeterminado ...

 Data Wrangler



 Copilot

Explorador

Elementos de datos

Recursos

Agregar elementos de datos

Items

LH_STORE_Bronce

ADN Fabric - Transformacion

Carga los archivos Parquet dentro de una carpeta en tablas del lakehouse BRONCE

- Utiliza el cuaderno NB_TRNSF_ParquetATabla
- El nombre de la tabla es el mismo nombre del archivo, sustituyendo el . por _

1

RUTA_CARPETA = 'Files/AdventureWorks'

[11]

✓ - Command executed in 357 ms by Fabric on 5:58:00 PM, 6/18/25

1

from pathlib import Path

2

3

for f in notebookutils.fs.ls(RUTA_CARPETA):

4

ruta_archivo_parquet = f.path

5

nombre_archivo = Path(f.name).stem

6

nombre_tabla = nombre_archivo.replace('.', '_')

7

print(f'Cargando el archivo {nombre_archivo} hacia la tabla {nombre_tabla}')

8

notebookutils.notebook.run('NB_TRNSF_ParquetATabla', 120, {'RUTAARCHIVO_PARQUET': ruta_archivo_parquet, 'NOMBRE_TABLA': nombre_ta

9

[12]

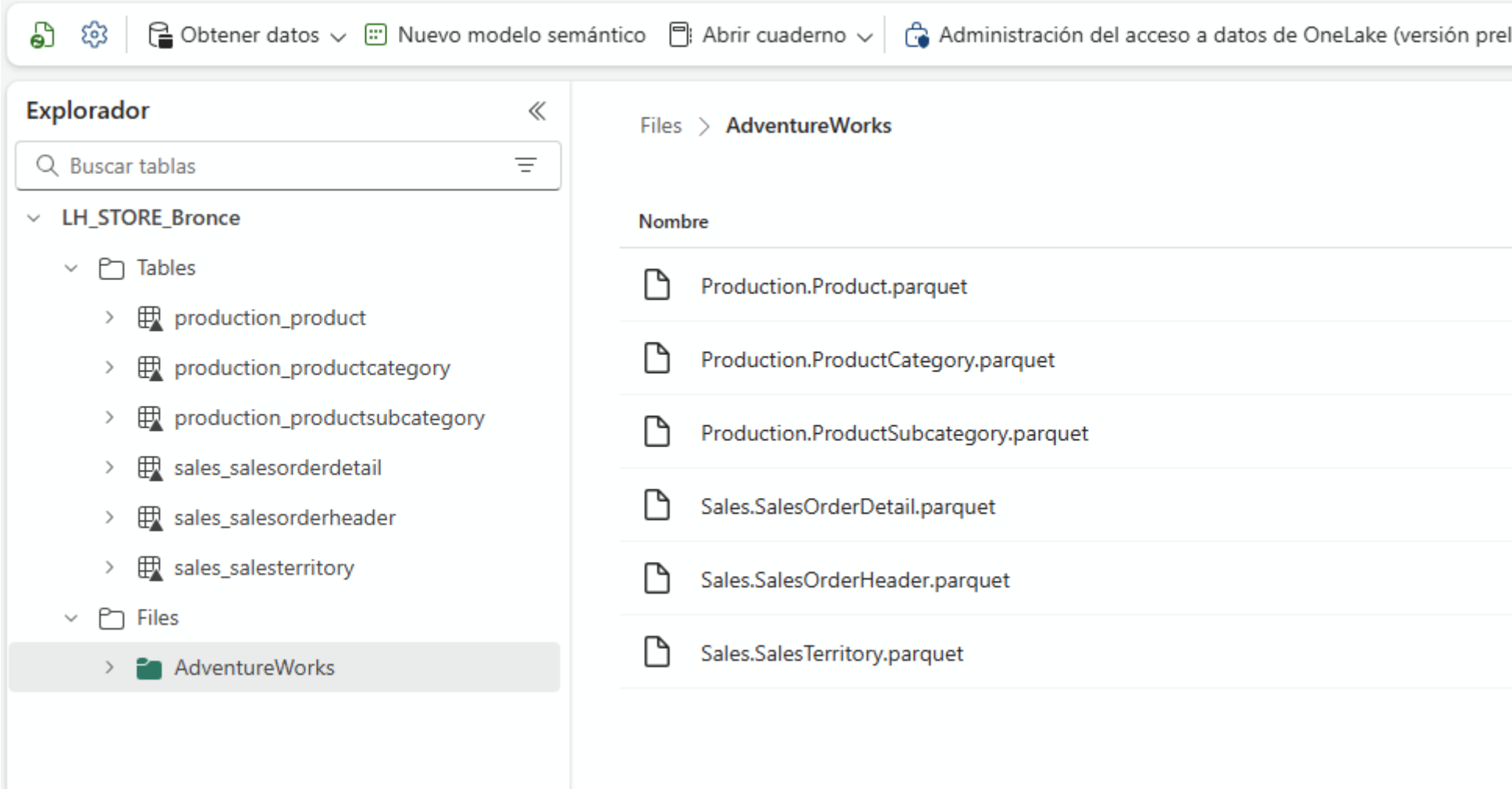
✓ - Command executed in 1 min 40 s 570 ms by Fabric on 5:59:41 PM, 6/18/25

Cargando el archivo Production.Product hacia la tabla Production_Product

Nombre de actividad	Instantánea	Estado	Progreso	Duración	Valor de :
0	NB_TRNSF_ParquetATabla	✓ Correcto	<div>100%</div>	17 s 424 ms	-

Capa Bronce

Tablas creadas en el Lakehouse



The screenshot displays the Microsoft Fabric Lakehouse Explorer interface. The top navigation bar includes icons for file operations and a menu with options: 'Obtener datos', 'Nuevo modelo semántico', 'Abrir cuaderno', and 'Administración del acceso a datos de OneLake (versión pre)'. The left sidebar, titled 'Explorador', contains a search bar 'Buscar tablas' and a tree view of the 'LH_STORE_Bronce' database. Under the 'Tables' folder, several tables are listed, including 'production_product', 'production_productcategory', 'production_productsubcategory', 'sales_salesorderdetail', 'sales_salesorderheader', and 'sales_salesterritory'. The 'Files' folder is also visible, with 'AdventureWorks' selected. The main pane on the right shows the 'Files > AdventureWorks' view, displaying a list of files with their names and file icons.

Nombre
Production.Product.parquet
Production.ProductCategory.parquet
Production.ProductSubcategory.parquet
Sales.SalesOrderDetail.parquet
Sales.SalesOrderHeader.parquet
Sales.SalesTerritory.parquet

Transformación con T-SQL

Capa Plata

A · D · N  Fabric

Transformaciones en la capa Plata

- Calidad de datos
- Limpieza
- Tipos de datos según sea el caso
- Eliminar columnas
- Combinar
 - Desnormalizar tablas para agrupar los atributos en las entidades
 - Datos geográficos
 - Datos por niveles, por ejemplo, Productos
 - Combinar cabecera y detalle

Capa Plata

Transformaciones con DF Spark

Transformación	Función PySpark
Elegir columnas	<code>df.select("column1", ..., "columnn")</code>
Renombrar columnas	<code>df.withColumnRenamed("Old Name", "New Name")</code>
Combinar consultas	<code>df.join(df_right, on="key", how="left right inner")</code>
Eliminar columnas	<code>df.drop("columna1", "columna2", ..., "columnn")</code>
Guardar datos en una tabla del LH	<code>df.write.mode("overwrite").format("delta").saveAsTable("LH_Name.TableName")</code>
Añadir o modificar columnas	<code>df. withColumn()</code>

Demos

Transformación con Flujos de Datos Gen 2 **Capa Plata**



Transformación con Flujos de Datos Gen2

Capa Oro

A · D · N  **Fabric**

Transformaciones en la capa Oro

- Filtrar por columnas
- Tipos de datos según sea el caso
 - Fecha vs Fecha hora
- Tabla Calendario
- Todas las tareas finales que requiera la creación y diseño del modelo semántico

Capa Oro

Funciones SQL (pyspark.sql.functions)

Transformación	Función PySpark
Trabajar con listas	<code>sequence, explode</code>
Indicar literal	<code>lit</code>
Funciones de fecha	<code>to_date, year, month, date_format, dayofmonth, quartes</code>

Capa Plata

Transformaciones con DF Pandas

Transformación	Función Python
Cambiar tipo a columnas	<code>df.astype({'Column1': int64 datetime64[ns] float64, 'Column2': int64 datetime64[ns] float64})</code>
Renombrar columnas	<code>df.withColumnRenamed("Old Name", "New Name")</code>
Eliminar columnas	<code>df.drop(columns=['Column1', 'Column2'])</code>

Demos

Transformación con Flujos de Datos Gen 2 **Capa Oro**





Siguiente encuentro

Orquestación

- Canalizaciones de datos
- Cuadernos Spark
- Air Flow

A · D · N  Fabric

¡Gracias!

A · D · N  **Fabric**

¿Preguntas?

Recursos - Spark

- <https://microsoftlearning.github.io/mslearn-fabric.es-es/Instructions/Labs/02-analyze-spark.html>
- <https://learn.microsoft.com/es-es/azure/databricks/getting-started/dataframes>
- <https://docs.databricks.com/aws/en/getting-started/dataframes>
- https://spark.apache.org/docs/3.5.5/api/python/getting_started/quickstart_df.html

Recursos – Data Wrangler

- <https://learn.microsoft.com/es-es/fabric/data-science/data-wrangler>
- <https://learn.microsoft.com/es-es/fabric/data-science/data-wrangler-spark>
- <https://marketplace.visualstudio.com/items?itemName=ms-toolsai.datawrangler>
- <https://learn.microsoft.com/es-es/fabric/data-science/data-wrangler>
- <https://learn.microsoft.com/es-es/training/modules/explore-data-for-data-science-microsoft-fabric/>

Recursos

- Blog oficial
 - <https://blog.fabric.microsoft.com/en-us/blog>
- Microsoft Learn Fabric en español
 - <https://learn.microsoft.com/es-es/fabric/>
- Curso en LinkedIn Learning en español / inglés
 - <https://www.linkedin.com/learning/search?entityType=COURSE&keywords=Fabric>
- Applied skills
 - https://learn.microsoft.com/en-us/credentials/browse/?credential_types=applied%20skills&products=fabric

Recursos



- Recursos - Microsoft
- <https://www.microsoft.com/en-us/microsoft-fabric/blog/2025/03/31/fabcon-2025-fueling-tomorrows-ai-with-new-agentic-capabilities-and-security-innovations-in-fabric/>
- <https://powerbi.microsoft.com/en-us/blog/grace-period-for-transitioning-from-power-bi-premium-to-microsoft-fabric/>

Recursos - Sitios, canales y blogs

- ADN Fabric
 - <https://www.linkedin.com/company/adnfabric>
- YouTube ADN Fabric
 - <https://www.youtube.com/@ADNFabric>
- Recursos de dataXbi
 - <https://www.dataxbi.com/fabric/>