




DP-700: Ingeniero de datos de Microsoft Fabric

Orquestación de procesos y movimiento de datos con Microsoft Fabric



Objetivos de aprendizaje

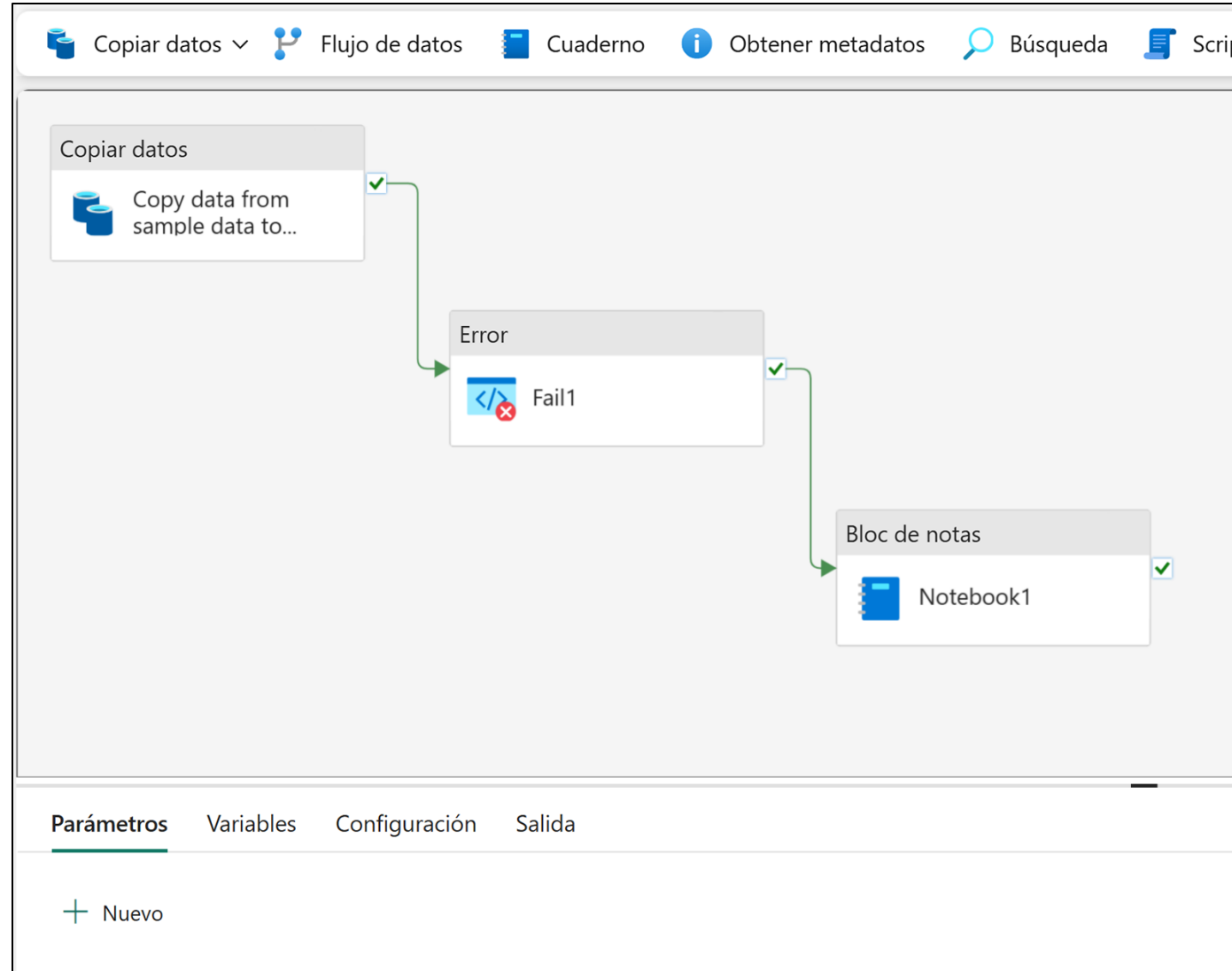


- Descripción de funcionalidades de pipeline en Microsoft Fabric.
- Uso de la actividad Copiar datos en un pipeline.
- Creación de pipelines basadas en plantillas predefinidas.
- Ejecución y supervisión de pipelines.

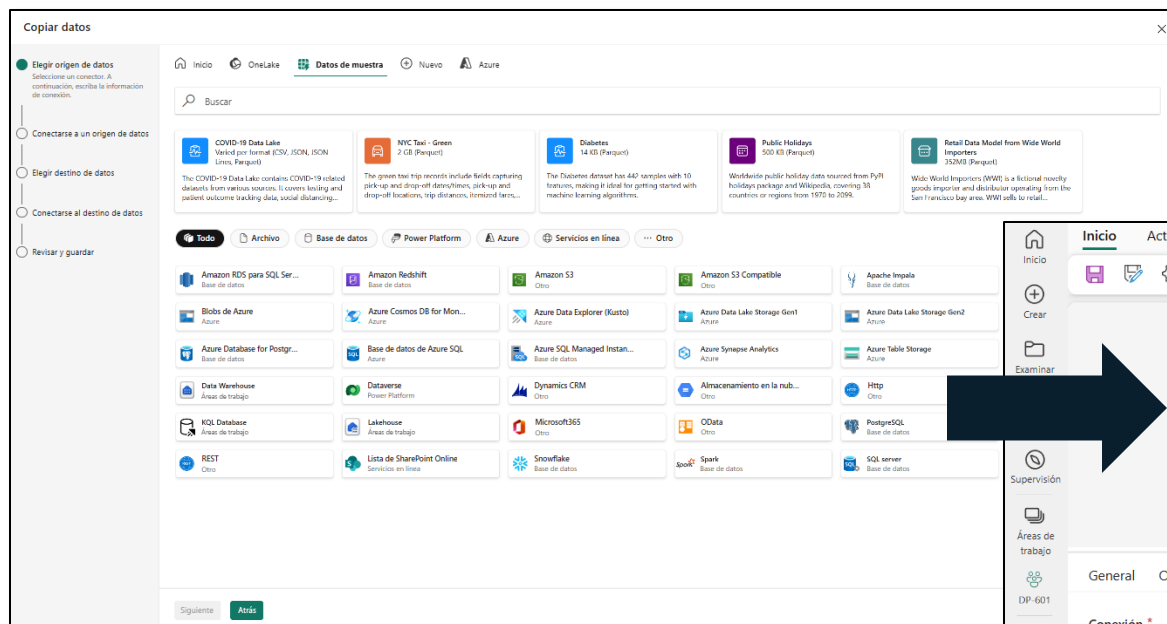
pipelines de Microsoft Fabric

Conceptos de la pipeline:

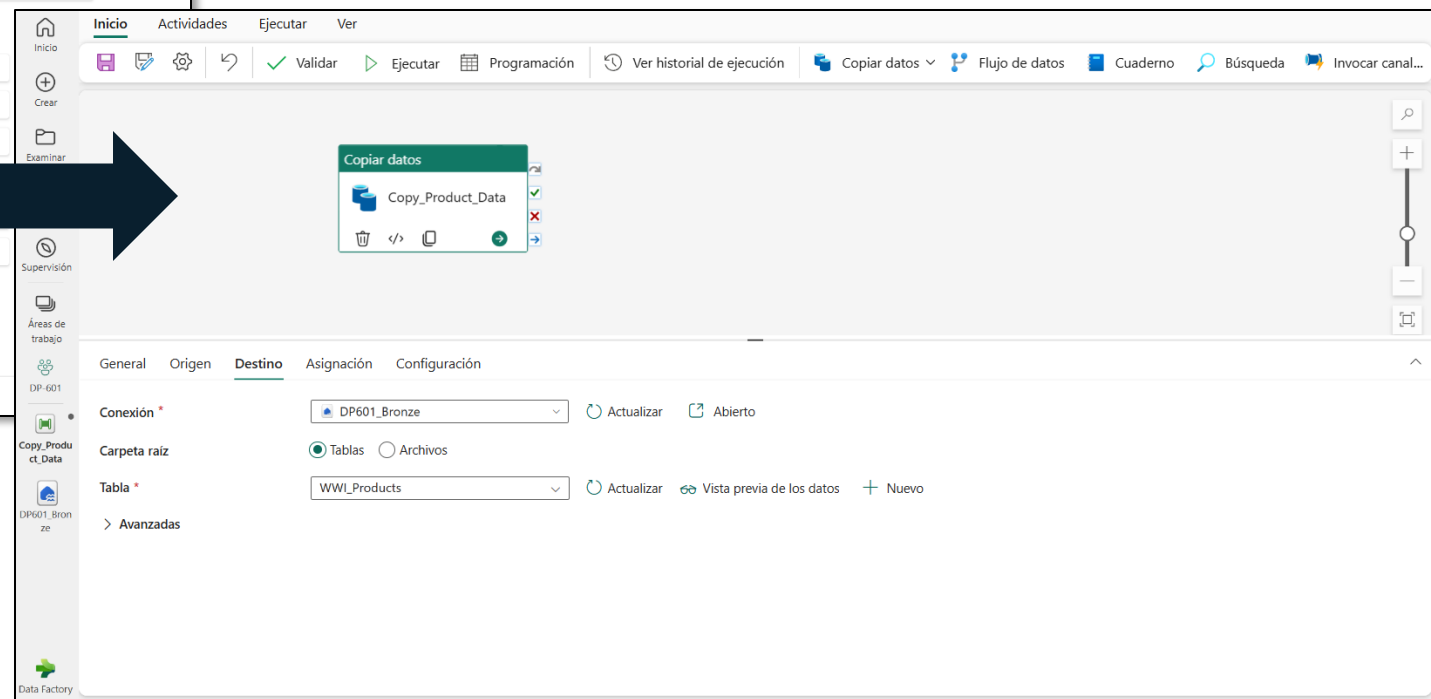
- Actividades
 - Transformación de datos
 - Flujo de control
- Parámetros
- Ejecuciones de la pipeline



Actividades comunes: Copiar datos



1. Uso de la herramienta de copia de datos.



2. Edición de la configuración debajo del lienzo de pipeline.

Uso de plantillas para actividades comunes



Copia masiva desde la base de datos

by Microsoft

Use this template to copy data in bulk from a database using an external control table to store the partition list of your source tables....



Copia masiva desde los archivos a la base de datos

by Microsoft

Use this template to copy data in bulk from Azure Data Lake Storage Gen2 to Azure SQL Database.
...



Copia de datos de ADLS Gen2 a un archivo Lakehouse

by Microsoft

Use this template to copy data from ADLS Gen2 to a specified file location in your Lakehouse.
...



Copia de datos de ADLS Gen2 a una tabla Lakehouse

by Microsoft

Use this template to copy data from ADLS Gen2 to a specified table in your Lakehouse.
...



Copia de datos de Azure SQL DB a una tabla de Lakehouse

by Microsoft

Use this template to copy data from your Azure SQL database to a specified table in your Lakehouse.
...



Copia de datos de datos de ejemplo en un archivo de...

by Microsoft

Use esta plantilla para copiar datos de datos de muestra (NYC Taxi - Green) en un archivo de almacenamiento de lago....



Copia de datos de datos de muestra en Warehouse

by Microsoft

Use esta plantilla para copiar datos de datos de ejemplo (Retail Data Model from Wide World Importers) a Fabric Warehouse....



Copia de varios contenedores de archivos entre almacenes d...

by Microsoft

Use this template to leverage multiple copy activities to copy containers or folders between file based stores, where each copy...

Ejecución y supervisión de pipelines

- 1. Validación
- 2. Ejecutar
- 3. Programación
- 4. Visualización del historial de ejecuciones

Inicio

Actividades

Ejecutar

Ver

Validar

Ejecutar

Programación

Ver historial de ejecución

Copiar datos

Copy from To process to In...

Eliminar datos

Delete di

General

Origen

Destino

Asignación

Configuración

Nombre *

Copy from To process to In process

Más información

Descripción

Tiempo de espera agotado

0.12:00:00

Reintentar

0

> Avanzadas

| Nombre de actividad | Estado | Hora de inicio |
|------------------------|----------|------------------|
| Process rides pipeline | Correcto | 3:17 PM, 6/21/23 |
| Process rides pipeline | Error | 2:31 PM, 6/28/23 |
| Process rides pipeline | Correcto | 2:35 PM, 6/28/23 |
| Process rides pipeline | Correcto | 2:44 PM, 6/28/23 |
| Process rides pipeline | Correcto | 2:50 PM, 6/28/23 |
| Process rides pipeline | Correcto | 3:41 PM, 6/28/23 |
| Process rides pipeline | Correcto | 3:50 PM, 6/28/23 |
| Process rides pipeline | Correcto | 3:33 PM, 6/30/23 |
| Process rides pipeline | Correcto | 3:55 PM, 6/30/23 |

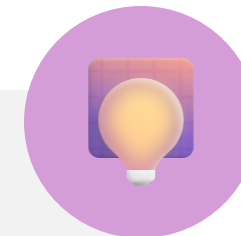
Ejercicio

60 minutos



Ingesta de datos con un pipeline

Prueba de conocimientos



1 ¿Qué es un pipeline de datos?

- ☐ Una carpeta especial en el almacenamiento de OneLake donde los datos se pueden exportar desde un almacén de lago.
- ☒ Una secuencia de actividades para orquestar un proceso de transformación o ingesta de datos.
- ☐ Una instancia de Power Query guardada.

2 Quieres usar un pipeline para copiar datos en una carpeta con un nombre especificado para cada ejecución. ¿Qué deberías hacer?

- ☐ Crear varias pipelines: una para cada nombre de carpeta.
- ☐ Usar un flujo de datos Gen2.
- ☒ Agregar un parámetro a la pipeline y usarlo para especificar el nombre de carpeta de cada ejecución.

3 Anteriormente, has ejecutado un pipeline que contiene varias actividades. ¿Cuál es la mejor forma de comprobar cuánto tiempo ha tardado en completarse cada actividad individual?

- ☐ Volver a ejecutar la pipeline y observar la salida, donde se asigna el tiempo de cada actividad.
- ☒ Ver los detalles del historial de ejecución correspondiente.
- ☐ Ver el valor actualizado para el modelo semántico predeterminado del almacén de lago de datos.

Resumen

En esta sección, se ha descrito lo siguiente:

- Cómo se pueden usar las pipelines para organizar procesos y movimiento de datos.
- Las actividades comunes y las plantillas listas para usar.
- La ejecución y la supervisión de pipelines.

Información adicional

Orquestación de procesos y movimiento de datos con
Microsoft Fabric

<https://aka.ms/fabric-pipeline-spa>



Organizar un almacén
de lago con arquitectura
de medallas

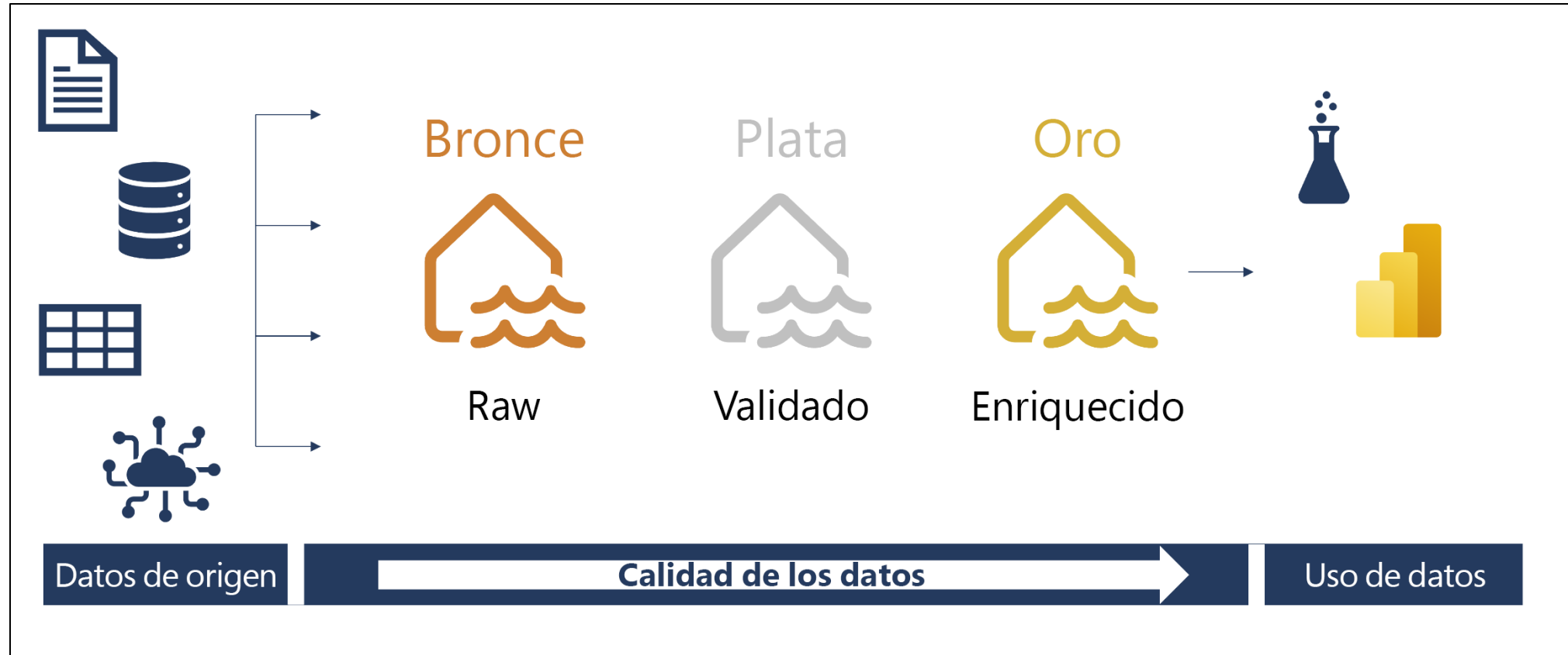


Objetivos de aprendizaje



- Describir los principios de uso de la arquitectura de medallas en la administración de datos.
- Aplicar el marco de arquitectura de medallas.
- Analizar los datos almacenados en el almacén de lago mediante DirectLake en Power BI.

Información general sobre la arquitectura de medallas



Movimiento y transformación de datos en los distintos niveles

- ¿Con cuántos datos estás trabajando?
- ¿Qué tan complejas son las transformaciones que necesitas realizar?
- ¿Con qué frecuencia necesitarás mover datos entre capas?
- ¿Con qué herramientas te sientes más cómodo?

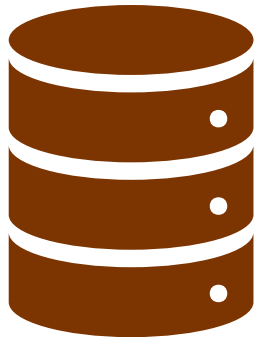
Nota:

Recuerda que los flujos de datos transforman los datos y las pipelines organizan los datos y los Notebooks pueden hacer las dos cosas.

Implementar una arquitectura de medallas

Bronce

Ingesta de datos sin procesar



pipelines, flujos de datos,
cuadernos

Plata

Limpieza y validación de datos



Flujos de datos o cuadernos

Oro

Transformaciones y modelado
adicionales



Punto de conexión de análisis
SQL o modelo semántico

Consultar e informar sobre los datos de Fabric Lakehouse

The screenshot displays the Microsoft Fabric Lakehouse interface. On the left, the 'Explorador' (Explorer) pane shows the 'Sales' database structure, including schemas, tables, and queries. The 'Sales' database is expanded, showing the 'dbo' schema with tables like 'dimcustomer...', 'dimdate_gold', 'dimproduct...', 'factsales_gold', and 'sales_silver'. The 'Queries' section is also visible, showing 'SQL query 1' and 'Visual query 1'.

The main area shows the 'SQL query 1' editor with the following query:

```
1 select CustomerName, SUM(Quantity) AS TotalQuantity
2 FROM [sales_silver]
3 GROUP BY CustomerName
4 ORDER BY TotalQuantity DESC
```

The 'Resultados' (Results) pane displays the query output as a table with 17 rows and 2 columns: 'CustomerName' and 'TotalQuantity'. The results are sorted by 'TotalQuantity' in descending order.

| | ABC CustomerName | 123 TotalQuantity |
|----|-------------------|-------------------|
| 1 | Samantha Jenkins | 41 |
| 2 | Henry Garcia | 39 |
| 3 | Charles Jackson | 38 |
| 4 | April Shan | 35 |
| 5 | Hailey Patterson | 34 |
| 6 | Ryan Thompson | 34 |
| 7 | Mason Roberts | 33 |
| 8 | Dalton Perez | 32 |
| 9 | Jason Griffin | 31 |
| 10 | Fernando Barnes | 28 |
| 11 | Jasmine Powell | 27 |
| 12 | Nicholas Brown | 25 |
| 13 | Ashley Henderson | 25 |
| 14 | Jennifer Simmons | 24 |
| 15 | Gina Martin | 24 |
| 16 | Ana Perry | 24 |
| 17 | Alexandra Jenkins | 22 |

The status bar at the bottom indicates 'Correcto (2 sec 671 ms)' and 'Columnas: 2 Filas: 10,000'.

Protección de las capas de medallas

**Control de acceso
a la capa Bronce:**
solo lectura.

**Utilización de
la capa Plata:**
compilación
opcional.

**Control de acceso
a la capa Oro:** solo
lectura.

Ejercicio

45 minutos



Organiza tu almacén de lago de Fabric utilizando una arquitectura de medallas

Prueba de conocimientos



- 1** ¿Cuál de los siguientes conjuntos de capas se asocia normalmente a la arquitectura de medallas para la administración de datos?
 - ☐ Sin procesar, pulido, refinado
 - ☒ Bronce, Plata, Oro
 - ☐ Inicial, Intermedio, Final
- 2** ¿Qué herramienta es más adecuada para la transformación de datos en Fabric cuando se trabaja con datos a gran escala que seguirán creciendo?
 - ☐ Dataflows Gen2
 - ☐ Pipelines
 - ☒ Blocs de notas
- 3** ¿Cuál es la ventaja de almacenar diferentes capas de tu almacén de lago de datos en áreas de trabajo independientes?
 - ☒ Puede mejorar la seguridad, administrar el uso de la capacidad y optimizar la rentabilidad.
 - ☐ Facilita el uso compartido de datos con compañeros.
 - ☐ No hay ninguna ventaja de almacenar diferentes capas de almacén de lago en áreas de trabajo independientes.

Resumen

En esta sección, se ha descrito lo siguiente:

- La implementación de una arquitectura de medallas con Microsoft Fabric.
- La consulta e información sobre los datos del almacén de lago de datos.
- La protección de capas de medallas.

Información adicional

Organizar un almacén de datos de Fabric mediante el diseño de la arquitectura de medallas


<https://aka.ms/fabric-medallion-spa>



Introducción a inteligencia en tiempo real en Microsoft Fabric



Objetivos de aprendizaje



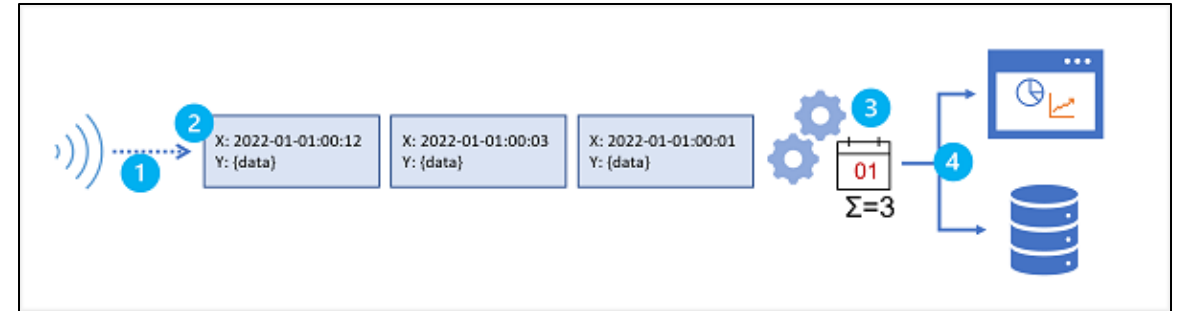
- Describir los conceptos de análisis de datos en tiempo real
- Explorar los componentes principales de la inteligencia en tiempo real en Microsoft Fabric

¿Qué es el análisis en tiempo real?

Ingesta y procesamiento de un *flujo* de datos perpetuo

Entre los objetivos comunes para el análisis en tiempo real se incluyen:

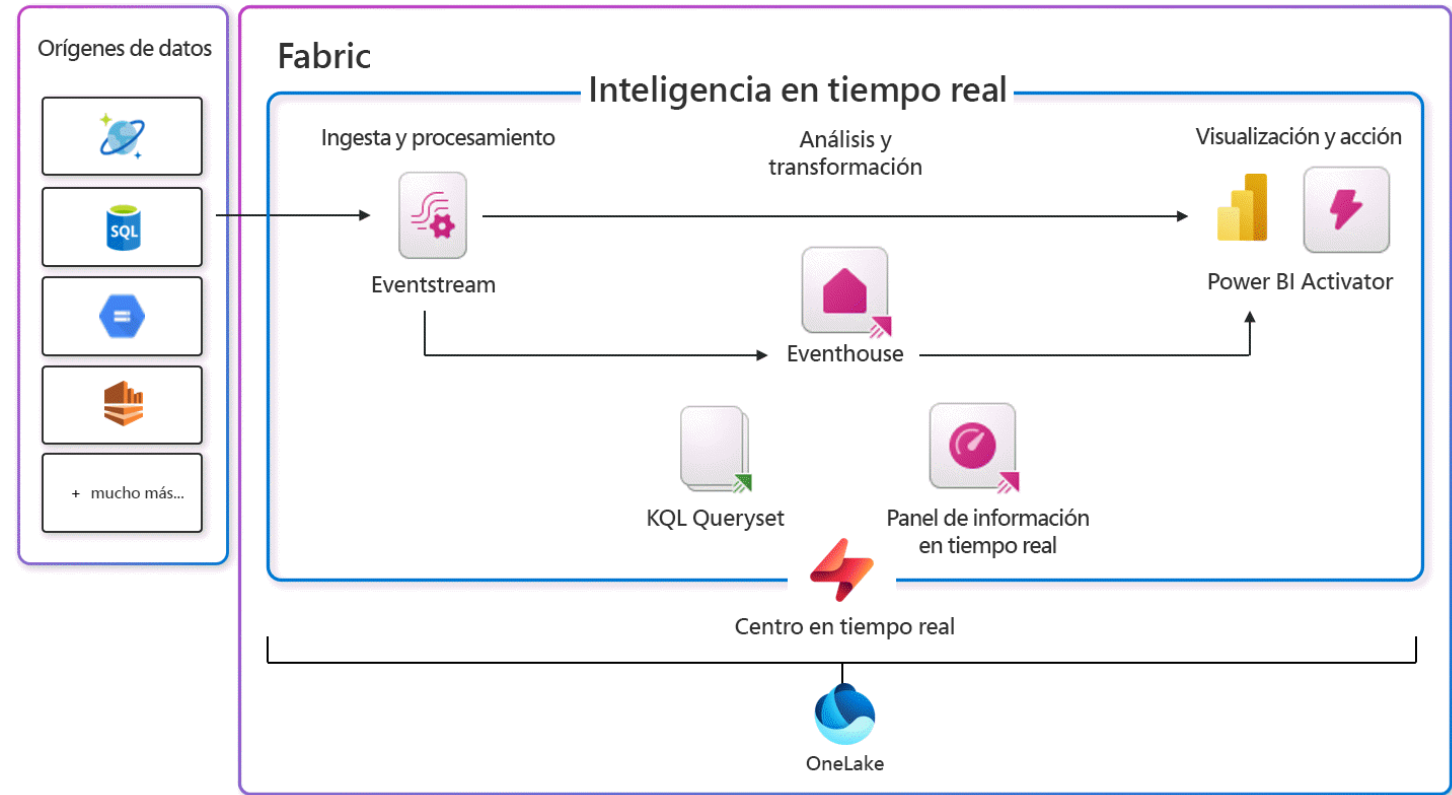
- Análisis continuo de los datos para informar sobre **problemas** o **tendencias**
- Descripción del comportamiento del componente o del sistema bajo varias condiciones para ayudar a planificar futuras mejoras
- Desencadenamiento de **acciones** o **alertas** específicas cuando se producen determinados **eventos** o se superan ciertos **umbrales**



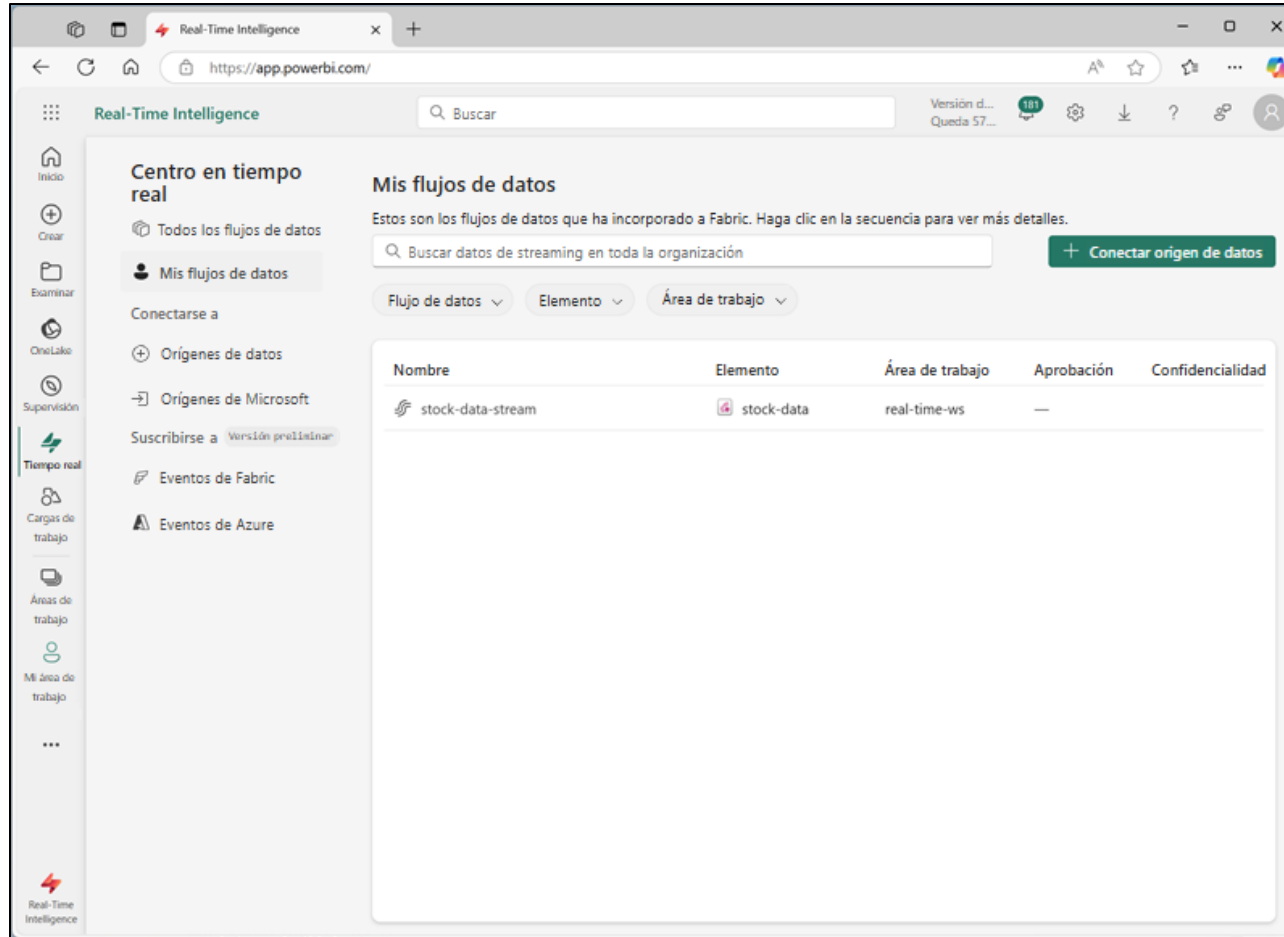
1. Un flujo de datos *no tiene límites*: los datos se añaden a la secuencia perpetuamente
2. Los datos normalmente incluyen datos *temporales* que indican cuándo se produjo el evento al que se relaciona el registro
3. La agregación de datos de streaming se realiza a menudo a través de *ventanas* temporales
4. Los resultados se pueden:
 - Usar para admitir la automatización o visualización en tiempo real (o *casi* en tiempo real)
 - Conservar en un almacén analítico que se va a combinar con otros datos para el análisis histórico
 - Ambos

Inteligencia en tiempo real en Microsoft Fabric

- **Eventstream**: ingesta y transformación de datos en tiempo real
- **Eventhouse**: almacenamiento escalable y consulta de datos en tiempo real
- **Conjunto de consultas de KQL**: herramienta para crear y ejecutar consultas en las bases de datos KQL dentro de un Eventhouse
- **Panel en tiempo real**: visualización interactiva y exploración de datos en tiempo real
- **Activator**: acciones automatizadas desencadenadas por eventos en tiempo real



Centro en tiempo real



- Buscar y conectarse a orígenes de datos en tiempo real y crear Eventstreams
- Suscribirse a eventos de Fabric y Azure y crear Eventstreams y alertas de Activator
- Obtener una vista previa y administrar las conexiones de datos en tiempo real, incluida la navegación a los datos de streaming capturados en un Eventhouse
- Crear paneles en tiempo real a partir de secuencias de eventos
- Aprobar y compartir recursos de datos en tiempo real en toda la organización

Ingesta y transformación de datos en tiempo real

Orígenes de datos:

- Servicios externos (Kafka, CDC,...)
- Eventos de Fabric
- Datos de ejemplo

Transformaciones de datos:

- Filtrado y administración de campos
- Agregado y agrupado
- Unión y combinación

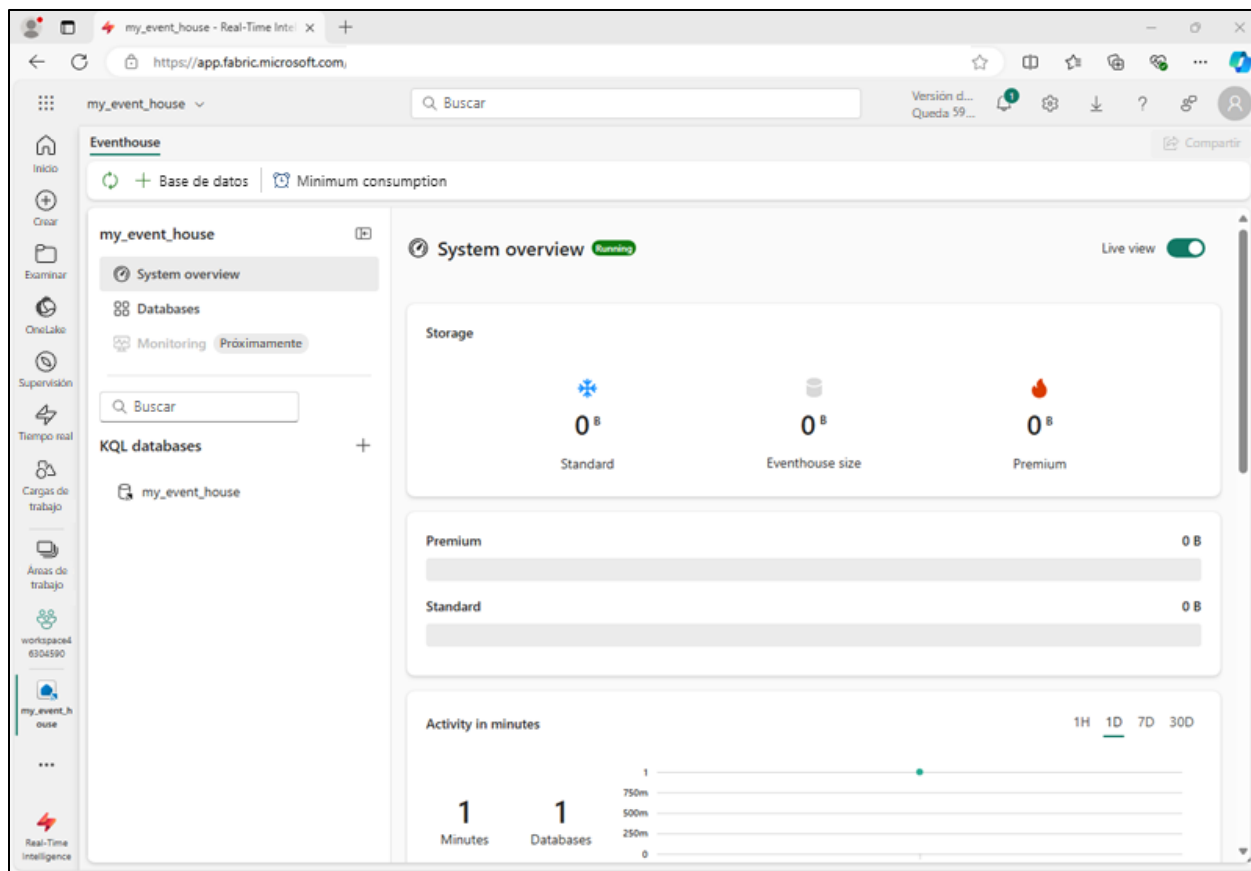
Destinos de datos:

- Eventhouse
- Lakehouse
- Secuencia derivada
- Fabric Activator
- Punto de conexión personalizado

The screenshot displays the Microsoft Fabric Real-Time Intelligence interface. The top navigation bar shows the URL <https://app.fabric.microsoft.com/>. The main workspace is titled 'stock-stream' and contains a data stream named 'stock-stream' with a transformation step labeled 'Transformar eventos o agregar destino'. The interface includes a sidebar with navigation options like 'Inicio', 'Crear', 'Examinar', 'OneLake', 'Supervisión', 'Áreas de trabajo', 'fab_workspace', 'stock-stream', 'my_eventhouse', and 'Real-Time Intelligence'. The bottom section shows the 'Resultado de la prueba' (Test Result) for the data stream, displaying a table of stock data.

| time | symbol | sector | securityType | bidPrice | bidSize |
|--------------------------|--------|--------------------|--------------|----------|---------|
| 2024-12-03T18:52:04.241Z | BMZM | retailing | commonstock | 2316.84 | 41 |
| 2024-12-03T18:52:04.241Z | NSFT | softwareservices | commonstock | 350.63 | 3 |
| 2024-12-03T18:52:04.241Z | HOQJ | mediaentertainment | commonstock | 1250.28 | 31 |
| 2024-12-03T18:52:04.241Z | BMZM | retailing | commonstock | 2306.84 | 123 |
| 2024-12-03T18:52:04.241Z | NSFT | softwareservices | commonstock | 370.63 | 23 |
| 2024-12-03T18:52:04.241Z | HOQJ | mediaentertainment | commonstock | 1370.28 | 66 |
| 2024-12-03T18:52:04.241Z | BMZM | retailing | commonstock | 2316.84 | 41 |
| 2024-12-03T18:52:04.241Z | NSFT | softwareservices | commonstock | 330.63 | 89 |
| 2024-12-03T18:52:04.241Z | HOQJ | mediaentertainment | commonstock | 1330.28 | 103 |

Almacenar y consultar datos en tiempo real

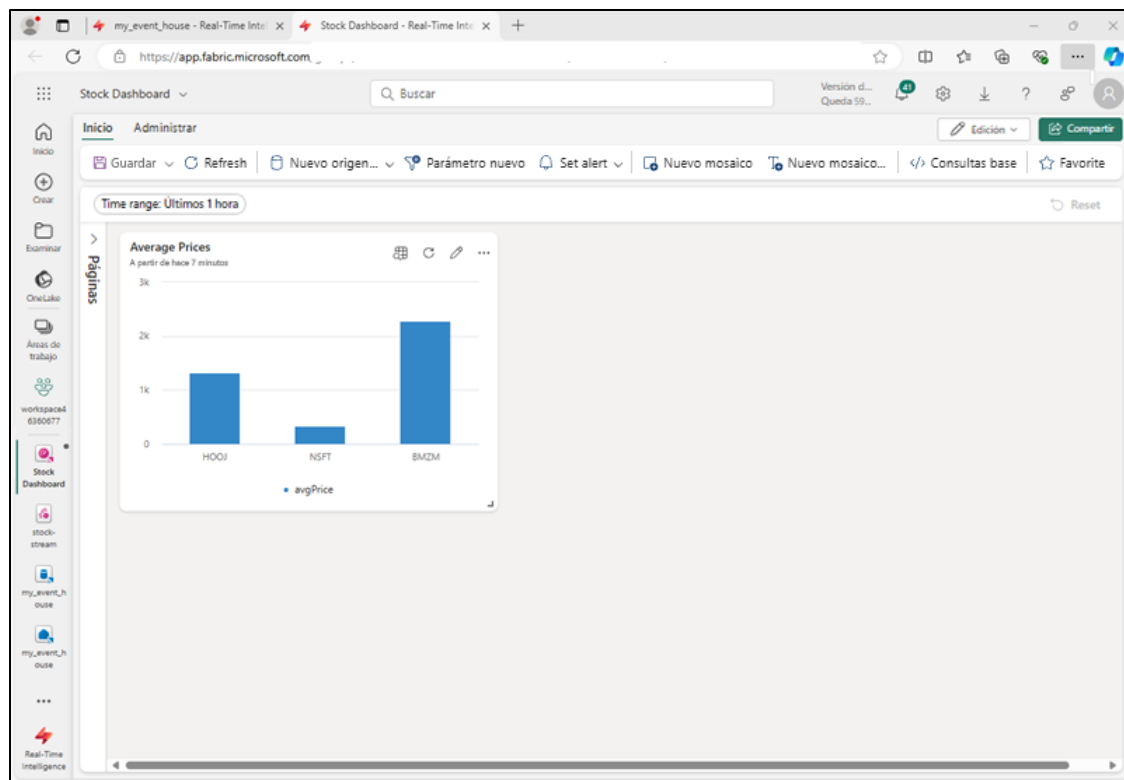


Eventhouse:

Almacenamiento escalable y consulta de datos en tiempo real en:

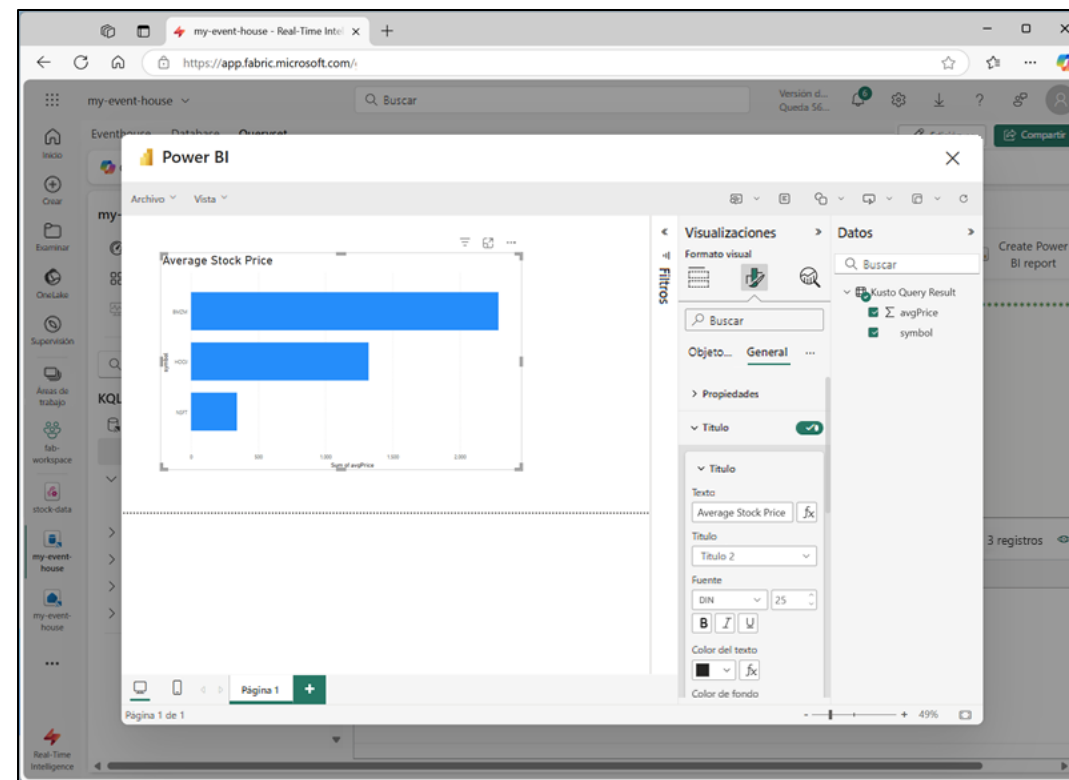
- **Bases de datos de KQL:**
- Almacenes de datos basados en el motor de base de datos de Kusto, optimizado para datos basados en el tiempo.
- Consultas con el Lenguaje de consulta Kusto (KQL), con consultas administradas en:
 - **Conjunto de consultas de KQL:**
 - Colecciones de consultas de KQL y SQL
 - Herramientas y plantillas integradas para tareas de consulta comunes
 - Copilot para la asistencia de la IA

Visualización de datos en tiempo real



Panel en tiempo real

- Un componente en la inteligencia en tiempo real de Fabric
- Iconos de visualización de datos interactivos basados en consultas de KQL



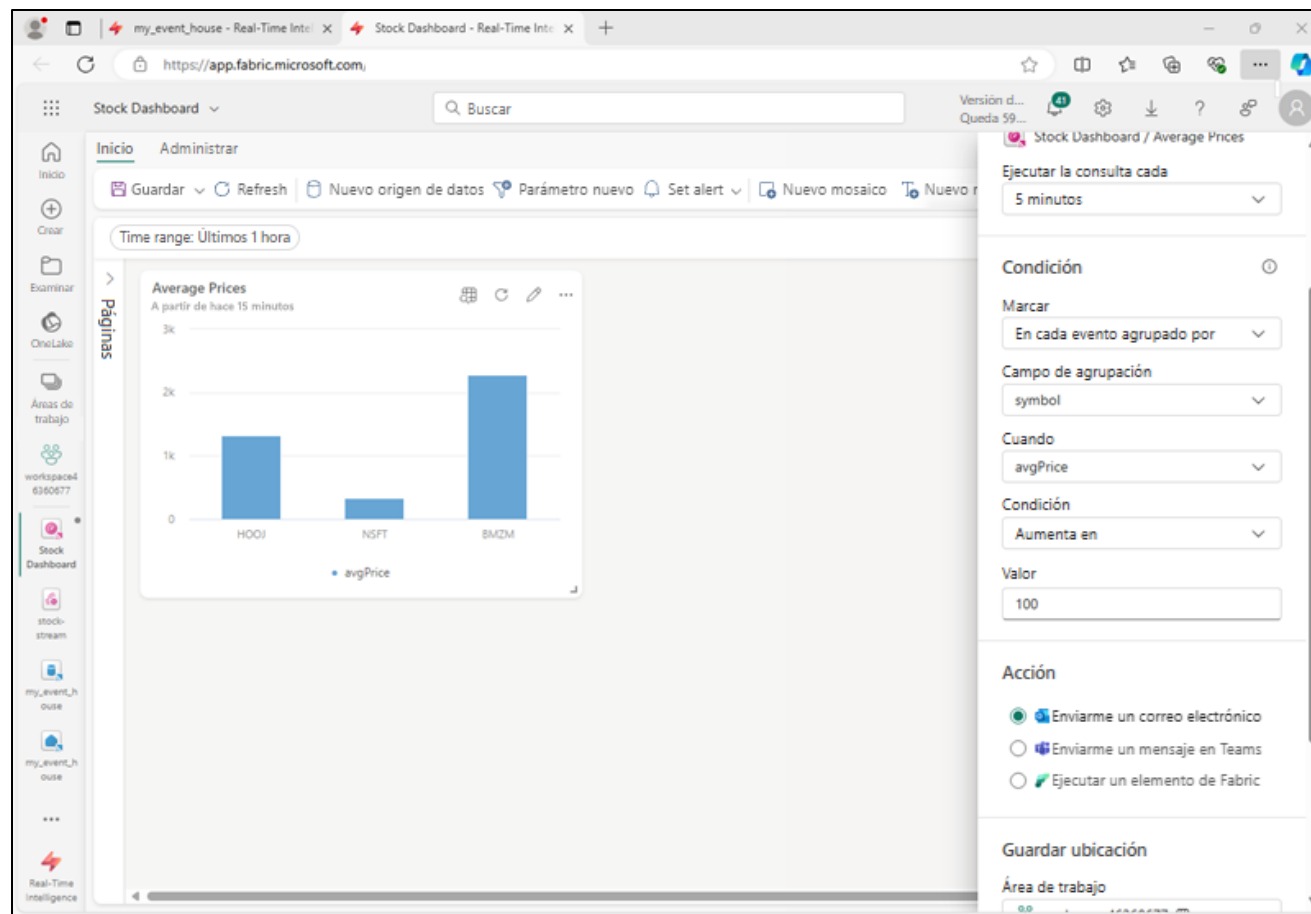
Power BI

- Componente principal de la plataforma Fabric
- Creación de informes de Power BI a partir de una base de datos KQL

Automatizar acciones

Uso de **Activator** para automatizar acciones:

- **Eventos:**
 - Cada registro de un flujo de datos representa un *evento*.
- **Objetos:**
 - Modelado de *objetos* conceptuales o reales basados en valores de datos
- **Reglas:**
 - Definición de *reglas* basadas en condiciones para desencadenar acciones
- **Propiedades:**
 - Uso de valores de datos de campo para definir *propiedades* de objeto y hacer referencia a ellos desde varias reglas



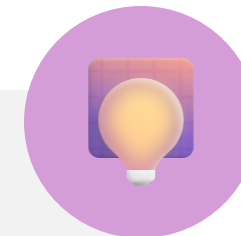
Ejercicio

30 minutos



Exploración de la inteligencia en tiempo real en Fabric

Prueba de conocimientos



1 ¿Qué componente de Inteligencia en tiempo real de Microsoft Fabric debe usar para ingerir y transformar un flujo de datos en tiempo real?

- ☐ Eventhouse
- ☒ Eventstream
- ☐ Activador

2 ¿Qué lenguaje está optimizado para consultar datos en tiempo real en un eventhouse?

- ☐ Python
- ☐ SQL
- ☒ KQL

3 ¿Qué componente de Inteligencia en tiempo real de Microsoft Fabric se usa para visualizar y explorar datos en tiempo real en iconos?

- ☐ Flujos de datos
- ☒ Paneles en tiempo real
- ☐ Tablas Delta

Resumen

En esta sección, se ha descrito lo siguiente:

- Exploración del Centro en tiempo real.
- Ingesta, transformación y carga de datos en tiempo real.
- Consulta y visualización de datos en tiempo real.
- Acciones de automatización.

Información adicional

Introducción a inteligencia en tiempo real en
Microsoft Fabric


<https://aka.ms/fabric-real-time-spa>



Uso de Eventstreams en tiempo real en Microsoft Fabric



Objetivos de aprendizaje



- Establecer el origen y los destinos en Eventstreams de Microsoft Fabric.
- Capturar, transformar y enrutar datos mediante Eventstreams de Microsoft Fabric.

Componentes de Eventstreams

Orígenes:

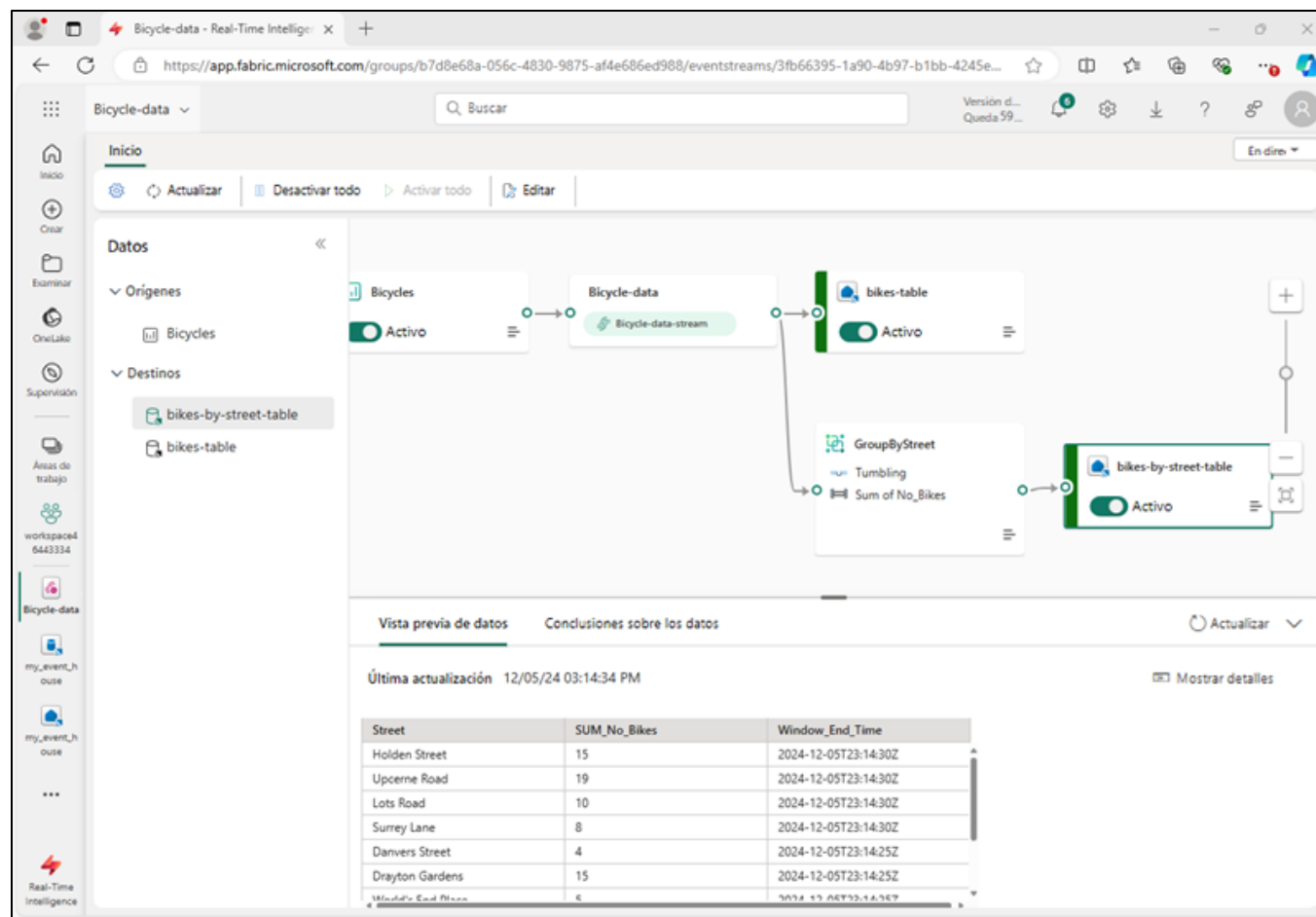
- De donde proceden los datos de streaming
- Incluye orígenes externos, eventos de Fabric y datos de ejemplo

Transformaciones:

- Operaciones que filtran o agregan los datos tal como se procesan desde el streaming
- Incluye funciones de ventanas temporales que permiten analizar eventos de datos dentro de períodos de tiempo discretos

Destinos:

- A donde se envían los datos de los eventos transformados
- Puede incluir Eventhouse, almacén de lago de datos, otro Eventstream o un Activator



Orígenes y destinos de Eventstream

| Sources | Destinos |
|--|---------------------------------|
| Centro de eventos de Azure | Eventhouse |
| Azure IoT Hub | Lakehouse |
| Captura de datos modificados de la base de datos | Punto de conexión personalizado |
| Azure SQL | Secuencia derivada |
| CDC de base de datos PostgreSQL: | Fabric Activator |
| CDC de base de datos MySQL | |
| CDC de Azure Cosmos DB | |
| Google Cloud Pub/Sub | |
| Flujos de datos de Amazon Kinesis | |
| Confluent Cloud Kafka | |
| Eventos de área de trabajo de Fabric | |
| Exploración de eventos de Azure Blob Storage | |
| Punto de conexión personalizado | |
| Datos de ejemplo | |

Transformaciones de Eventstream

Transformaciones

- Filtro
- Agregar
- Agrupar por
- Unión
- Expandir
- Join

Ventanas temporales

- Ventana de saltos de tamaño constante:
Divide los eventos entrantes en intervalos fijos que no se superponen
- Ventana deslizante:
Incluye eventos en intervalos fijos y superpuestos en función del tiempo
- Ventana de sesión:
Divide los eventos en intervalos variables y no superpuestos en función de las brechas entre las actividades
- Ventana de salto:
Divide los eventos entrantes en intervalos superpuestos fijos
- Ventana de instantánea:
Agrupar eventos de Eventstream que tienen la misma marca de tiempo

Ejercicio



30 minutos

**Ingesta de datos en tiempo real con
Eventstream de Microsoft Fabric**

Prueba de conocimientos



1 ¿Cuál es la función principal de Microsoft Fabric Eventstreams?

- ☐ Realizar la carga por lotes de datos en un almacenamiento de datos
- ☐ Ejecutar Notebooks y trabajos de Apache Spark
- ☒ Ingerir y transformar datos en tiempo real

2 ¿Qué permite hacer la transformación de ventana temporal?

- ☒ Agregar datos de eventos en un flujo en función de períodos de tiempo específicos
- ☐ Habilitar y deshabilitar automáticamente el streaming en horas programadas
- ☐ Eliminar los datos expirados después de un período de retención especificado

3 ¿Cuál es el propósito del destino de *Fabric Activator* en un Eventstream?

- ☒ Los datos enviados a un destino de Activator se pueden usar para desencadenar una acción automatizada basada en valores de datos
- ☐ Los datos ingeridos en un almacén de lago de datos o un Eventhouse no son visibles hasta que los activa Fabric Activator
- ☐ Activator restablece cualquier transformación que se haya aplicado en el Eventstream y devuelve los datos a un estado limpio

Resumen

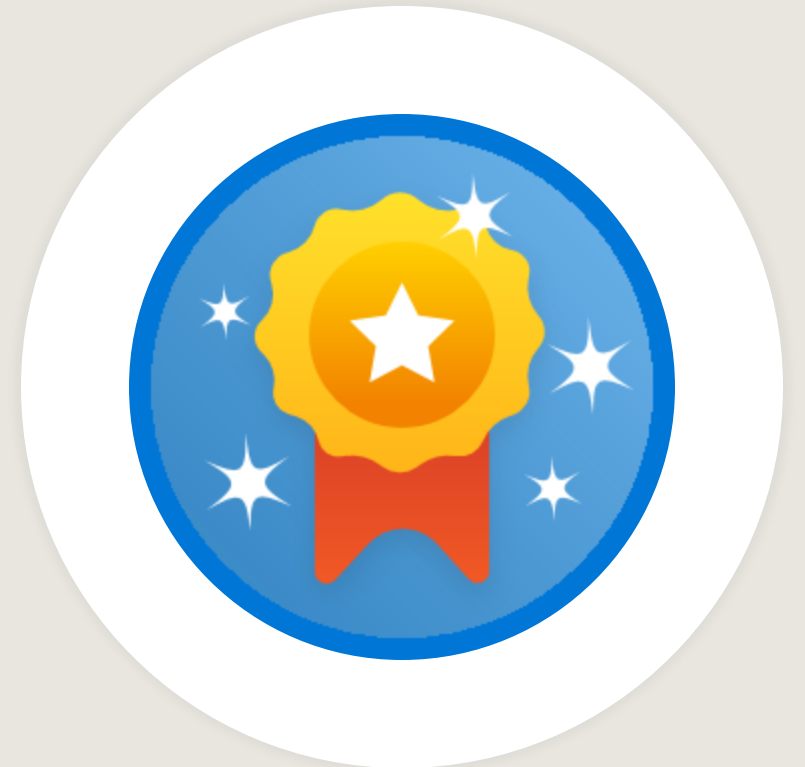
En esta sección, se ha descrito lo siguiente:

- Los componentes de Eventstream.
- Los orígenes y los destinos de Eventstreams.
- Las transformaciones y las características de las ventanas temporales.

Información adicional

Uso de secuencias de eventos en tiempo real en Microsoft Fabric

<https://aka.ms/mslearn-eventstreams-fabric-spa>



Trabajar con datos en tiempo real en un centro de eventos de Microsoft Fabric



Objetivos de aprendizaje



- Creación de un centro de eventos en Microsoft Fabric
- Consulta de datos en tiempo real mediante el Lenguaje de consulta Kusto (KQL)
- Creación de vistas materializadas y funciones almacenadas en una base de datos KQL

Introducción a un Eventhouse

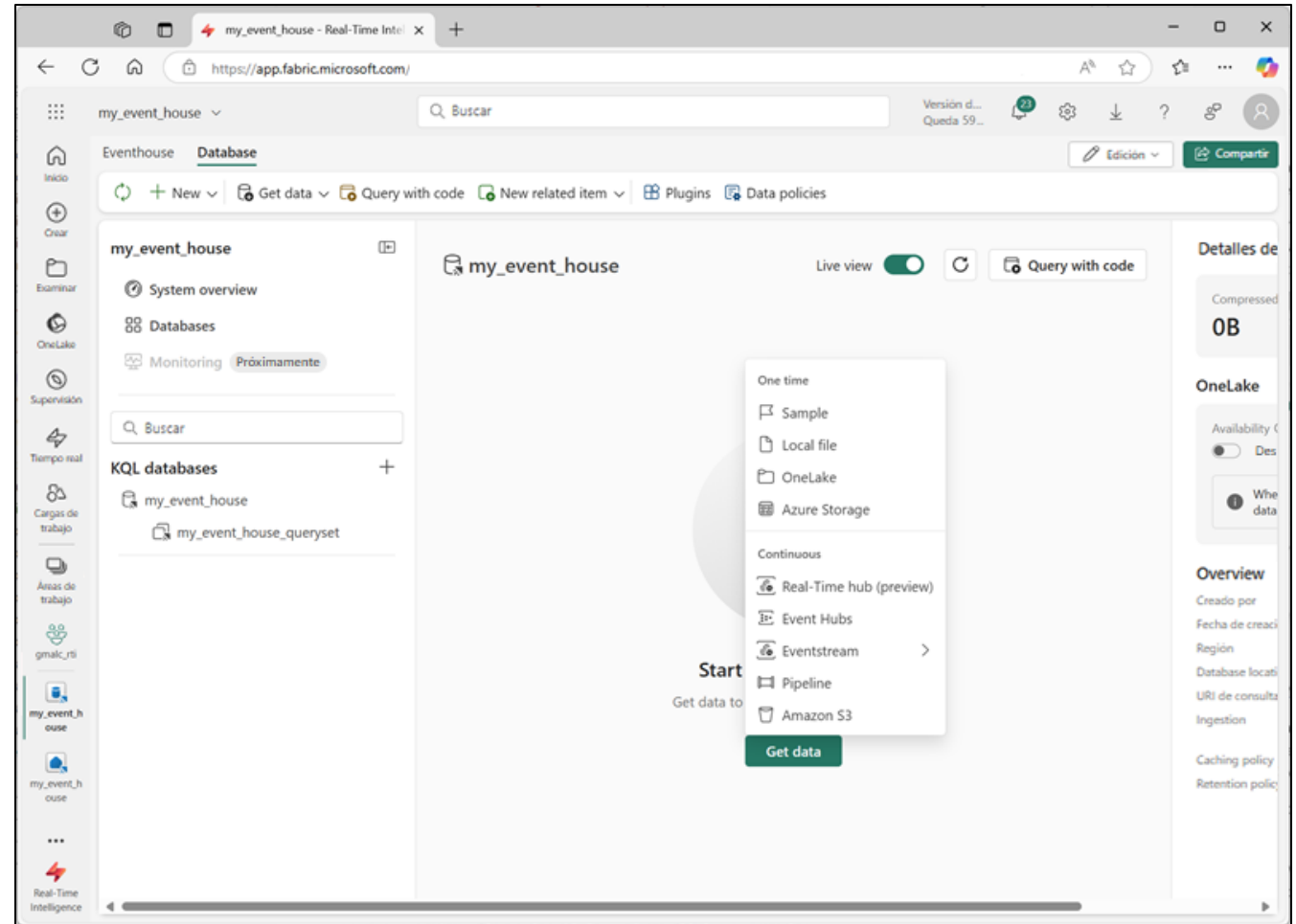
- Un Eventhouse puede contener una o varias bases de datos KQL
- Importación de datos en tablas desde orígenes de datos estáticos o de streaming
- Consulta de tablas mediante KQL:

Automotive

```
| where fare_amount > 20  
| project trip_id, pickup_datetime, fare_amount  
| sort by pickup_datetime desc
```

- O SQL:

```
SELECT trip_id, pickup_datetime, fare_amount  
FROM Automotive  
WHERE fare_amount > 20  
ORDER BY pickup_datetime DESC
```



Uso eficaz de KQL

Columnas de filtro

```
Automotive  
| project trip_id, vendor_id, pickup_datetime, fare_amount
```

Filtrar filas

```
Automotive  
| where pickup_datetime > ago(30min)  
| project trip_id, vendor_id, pickup_datetime, fare_amount
```

Resumen de datos

```
Automotive  
| where ingestion_time() > ago(1d)  
| summarize average_fare = avg(fare_amount) by vendor_id, pickup_hour = hourofday(pickup_datetime)  
| project pickup_hour, vendor_id, average_fare  
| sort by pickup_hour
```


Vistas materializadas y funciones almacenadas

Vistas materializadas

```
.create async materialized-view with (backfill=true) TripsByVendor on table Automotive  
{ Automotive  
  | summarize trips = count() by vendor_id, pickup_date = format_datetime(pickup_datetime, "yyyy-MM-dd")  
}
```

```
TripsByVendor  
| project pickup_date, vendor_id, trips  
| sort by pickup_date desc
```

Funciones almacenadas

```
.create-or-alter function trips_by_min_passenger_count(num_passengers:long)  
{ Automotive  
  | where passenger_count >= num_passengers  
  | project trip_id, pickup_datetime  
}
```

```
trips_by_min_passenger_count (3)  
| take 10
```

Ejercicio



30 minutos

Trabajo con los datos de un centro de eventos de Microsoft Fabric

Prueba de conocimientos



1 ¿Qué es un Eventhouse?

- ☐ Una tabla que realiza un seguimiento de los eventos de carga de datos en un almacenamiento de datos
- ☒ Un almacén de datos para datos en tiempo real en bases de datos KQL
- ☐ Otro nombre para un almacén de lago de datos en Microsoft Fabric

2 ¿Cuál es el propósito de usar el operador *project* en las consultas KQL (Lenguaje de consulta Kusto)?

- ☐ Aplicar agregaciones como *contar*, *suma* y *promedio*
- ☒ Especificar qué columnas se van a incluir en la salida de la consulta
- ☐ Filtrar los datos basados en condiciones específicas

3 ¿Cómo puedes crear una consulta parametrizada reutilizable para una base de datos KQL?

- ☒ Crear una función almacenada
- ☐ Crear una vista materializada
- ☐ Escribir un script de Python personalizado y guardarlo como una definición de trabajo de Spark

Información adicional

Trabajar con datos en tiempo real en un centro de eventos de Microsoft Fabric

<https://aka.ms/mslearn-eventhouse-fabric-spa>

