



**Escuela Superior  
de Ingeniería y Tecnología**  
Universidad de La Laguna

# Trabajo de Fin de Grado

Grado en Ingeniería Informática

---

## Desarrollo de una Plataforma de Datos en Microsoft Fabric

*Development of a Data Platform in Microsoft Fabric*

Francisco Javier García González

---

**La Laguna, 11 de julio de 2024**

**D. José Luis Roda García**, profesor Titular de Universidad adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor

**D. Ginés León Rodríguez**, responsable del departamento de Big Data & Data Science de Transporte Interurbanos de Tenerife S.A (T.I.T.S.A), como cotutor

## **C E R T I F I C A N**

Que la presente memoria titulada:

*“Desarrollo de una Plataforma de Datos en Microsoft Fabric”*

ha sido realizada bajo su dirección por D. Francisco Javier García González.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 11 de julio de 2024

# Agradecimientos

En primer lugar, agradecer a mi tutor, José Luis Roda García, y cotutor, Ginés León Rodríguez, por cederme su tiempo, conocimiento y ayuda a lo largo de este proyecto.

También agradecer a Andrea Di Teodoro, Ingeniero experto en Microsoft Fabric, y a Melanie Calero Sanchez, por su orientación durante el proyecto.

Gracias a los docentes del Grado en Ingeniería Informática que, con su esfuerzo y dedicación me han aportado los conocimientos y habilidades necesarias para avanzar en el grado.

Y por último, pero no por ello menos importante, agradecer a mis amigos y familiares, en especial a mi padre y a mi madre, quienes han sido un pilar fundamental a la hora de seguir adelante durante mi etapa universitaria.

# Licencia



© Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.

## Resumen

*El objetivo de este trabajo ha sido desarrollar un prototipo funcional para el análisis del tráfico en el Macizo de Anaga utilizando datos simulados. La plataforma de datos utilizada ha sido Microsoft Fabric, aprovechando su capacidad para la recolección, almacenamiento y procesamiento de grandes volúmenes de datos en un Lakehouse.*

*En primer lugar, se ha configurado un sistema completo en Microsoft Fabric, que incluye la recolección y almacenamiento de datos en un Lakehouse, así como la creación de canalizaciones de datos para automatizar el flujo de trabajo. Este sistema permite gestionar datos estructurados y no estructurados, optimizando el almacenamiento y procesamiento de la información necesaria para el análisis del tráfico.*

*Posteriormente, se ha utilizado la herramienta Power BI para la visualización de los resultados. A través de informes detallados, se pueden identificar patrones y tendencias en la movilidad y tráfico en la zona de Anaga. Power BI es fundamental para ofrecer una representación clara y accesible de los datos, facilitando la toma de decisiones basadas en información precisa.*

**Palabras clave:** Plataforma de datos, Lakehouse, Microsoft Fabric, Movilidad, Prototipo, Power BI.

## **Abstract**

*The objective of this work has been to develop a functional prototype for traffic analysis in the Anaga massif using simulated data. The data platform used has been Microsoft Fabric, taking advantage of its capacity for the collection, storage and processing of large volumes of data in a Lakehouse.*

*First, a complete system has been configured in Microsoft Fabric, which includes the collection and storage of data in a Lakehouse, as well as the creation of data pipelines to automate the workflow. This system allows the management of structured and unstructured data, optimizing the storage and processing of the information required for traffic analysis.*

*Subsequently, the Power BI tool was used to visualize the results. Through detailed reports, patterns and trends in mobility and traffic in the Anaga area can be identified. Power BI is essential to provide a clear and accessible representation of the data, facilitating decision making based on accurate information.*

**Keywords:** Data Platform, Lakehouse, Microsoft Fabric, Mobility, Prototype, Power BI

# Índice general

|  |           |
|--|-----------|
| <b>Capítulo 1 Introducción.....</b>              | <b>11</b> |
| 1.1 Antecedentes.....                            | 11        |
| 1.2 Desafíos.....                                | 11        |
| <b>Capítulo 2 Plataformas de datos.....</b>      | <b>13</b> |
| 2.1 Definición y requisitos.....                 | 13        |
| 2.2 Microsoft Azure.....                         | 14        |
| 2.3 Amazon Web Services.....                     | 15        |
| 2.4 Google Cloud Platform.....                   | 15        |
| 2.5 Conclusiones.....                            | 16        |
| <b>Capítulo 3 Microsoft Fabric.....</b>          | <b>17</b> |
| 3.1 Definición.....                              | 17        |
| 3.2 Power BI.....                                | 18        |
| 3.3 Data Factory.....                            | 19        |
| 3.3.1 Flujo de datos Gen2.....                   | 20        |
| 3.3.2 Canalizaciones de datos.....               | 21        |
| 3.4 Data Activator.....                          | 22        |
| 3.4.1 Eventos.....                               | 22        |
| 3.4.2 Objetos.....                               | 22        |
| 3.4.3 Desencadenadores.....                      | 23        |
| 3.4.4 Propiedades.....                           | 23        |
| 3.5 Industry solutions.....                      | 23        |
| 3.6 Real-Time Intelligence.....                  | 23        |
| 3.6.1 Detección de datos de streaming.....       | 25        |
| 3.6.2 Conexión a datos de streaming.....         | 25        |
| 3.6.3 Procesar flujos de datos.....              | 25        |
| 3.6.4 Almacenar y analizar datos.....            | 25        |
| 3.6.5 Visualización de información de datos..... | 26        |
| 3.6.6 Acciones de desencadenador.....            | 26        |
| 3.7 Data Engineering.....                        | 26        |
| 3.7.1 Lakehouse.....                             | 27        |
| 3.7.2 Notebook.....                              | 28        |
| 3.7.3 Entorno.....                               | 28        |
| 3.7.4 Definición de trabajo de Apache Spark..... | 28        |

|   |           |
|---|-----------|
| 3.8 Data Science.....                                     | 28        |
| 3.9 Data Warehouse.....                                   | 29        |
| 3.10 OneLake.....   | 30        |
| 3.11 Flujos de tareas en Microsoft Fabric.....            | 32        |
| <b>Capítulo 4 Prototipo implementado.....</b>             | <b>33</b> |
| 4.1 Crear un área de trabajo.....                         | 33        |
| 4.2 Crear un Lakehouse.....                               | 34        |
| 4.3 Crear una canalización de datos.....                  | 34        |
| 4.4 Crear un flujo de datos.....                          | 36        |
| 4.5 Crear un modelo semántico.....                        | 39        |
| 4.6 Crear un informe en Power BI.....                     | 40        |
| 4.7 Monitorización de métricas de uso.....                | 45        |
| 4.8 Diagrama de tareas.....                               | 46        |
| <b>Capítulo 5 Conclusiones y líneas futuras.....</b>      | <b>47</b> |
| 5.1 Conclusiones.....                                     | 47        |
| 5.2 Líneas futuras.....                                   | 47        |
| <b>Capítulo 6 Conclusions and future development.....</b> | <b>48</b> |
| 6.1 Conclusions.....                                      | 48        |
| 6.2 Future development.....                               | 48        |
| <b>Capítulo 7 Presupuesto.....</b>                        | <b>49</b> |

# Índice de tablas

**Tabla 7.1:** Coste de desarrollo

49

# Índice de figuras

|   |    |
|---|----|
| <b>Figura 3.1:</b> Arquitectura de Microsoft Fabric   | 16 |
| <b>Figura 3.2:</b> Componentes de Microsoft Fabric  | 17 |
| <b>Figura 3.3:</b> Funcionalidades de Data Factory.   | 19 |
| <b>Figura 3.4:</b> Ejemplo de un flujo de datos   | 19 |
| <b>Figura 3.5:</b> Pantalla de opción para obtener orígenes de datos                              | 20 |
| <b>Figura 3.6:</b> Ejemplo de canalizador de datos  | 21 |
| <b>Figura 3.7:</b> Diagrama de la arquitectura de inteligencia en tiempo real en Microsoft Fabric | 23 |
| <b>Figura 3.8:</b> Funcionalidades de Real-Time Intelligence                                      | 24 |
| <b>Figura 3.9:</b> Funcionalidades de Data Engineering  | 25 |
| <b>Figura 3.10:</b> Ejemplo del explorador de Lakehouse   | 26 |
| <b>Figura 3.11:</b> Funcionalidades de Data Science   | 27 |
| <b>Figura 3.12:</b> Funcionalidades de Data Warehouse   | 29 |
| <b>Figura 3.13:</b> Jerarquía de un inquilino en Microsoft Fabric                                 | 30 |
| <b>Figura 3.14:</b> Arquitectura de OneLake   | 30 |
| <b>Figura 3.15:</b> Ejemplo de diagrama de tareas   | 31 |
| <b>Figura 4.1:</b> Interfaz de las áreas de trabajo en Microsoft Fabric                           | 33 |
| <b>Figura 4.2:</b> Configuración de origen de datos   | 34 |
| <b>Figura 4.3:</b> Configuración de destino de datos  | 34 |
| <b>Figura 4.4:</b> Canalizador de datos con la copia de datos                                     | 35 |
| <b>Figura 4.5:</b> Obtener datos desde un flujo de datos Gen2                                     | 35 |
| <b>Figura 4.6:</b> Cambio de tipo de la columna   | 36 |
| <b>Figura 4.7:</b> Script en M que combina las columnas Fecha y Hora                              | 36 |
| <b>Figura 4.8:</b> Pasos aplicados dentro del flujo de datos                                      | 37 |
| <b>Figura 4.9:</b> Flujo de datos configurado dentro de la canalización                           | 38 |
| <b>Figura 4.10:</b> Creación de un modelo semántico   | 38 |
| <b>Figura 4.11:</b> Modelo semántico configurado en la canalización de datos                      | 39 |
| <b>Figura 4.12:</b> Selección de origen de datos desde Power BI desktop                           | 40 |
| <b>Figura 4.13:</b> Pestaña de menú   | 41 |
| <b>Figura 4.14:</b> Pestaña de agrupación de vehículos de entrada por fecha y hora                | 42 |
| <b>Figura 4.15:</b> Pestaña de tráfico de vehículos por zonas de entrada                          | 43 |
| <b>Figura 4.16:</b> Pestaña de tráfico de vehículos por días de la semana                         | 44 |
| <b>Figura 4.17:</b> Informe de métricas de uso  | 45 |
| <b>Figura 4.18:</b> Flujo de tareas del área de trabajo en Microsoft Fabric                       | 45 |

# Capítulo 1 Introducción

En la actualidad, el turismo emergente ha desempeñado un papel significativo en la transformación de numerosas áreas urbanas y rurales, generando tanto oportunidades como desafíos para las comunidades locales. De forma general, en zonas turísticas, el crecimiento del turismo ha resultado en un aumento considerable de tráfico de vehículos, lo que ha intensificado los problemas de movilidad. Este incremento de visitantes ha puesto de manifiesto la necesidad de implementar soluciones innovadoras que no sólo aborden los desafíos de movilidad, sino que también preserven y mejoren la calidad de vida de los residentes locales.

La presente memoria de Trabajo de Fin de Grado se centra en la evaluación de plataformas de datos para la gestión inteligente de la movilidad ante el exceso de tráfico en el macizo de Anaga. Se propone el desarrollo de un prototipo de proyecto que integre tecnologías avanzadas y estrategias de movilidad efectivas para hacer frente al problema de tráfico y movilidad que afecta tanto a los residentes como a los visitantes de la zona.

El objetivo principal es definir una estrategia de movilidad integral que aproveche las tecnologías emergentes y los datos recopilados mediante dispositivos de monitoreo para informar la toma de decisiones en tiempo real. Esto implica la recopilación y análisis de datos de tráfico, comportamiento del transporte público, patrones de movilidad de los usuarios, entre otros, que serán procesados y visualizados a través de herramientas de Business Intelligence (BI) para proporcionar información a los responsables de la planificación urbana y la gestión del transporte.

A través de este trabajo, se busca contribuir al desarrollo de soluciones innovadoras y sostenibles para mejorar la calidad de vida en la zona del macizo de Anaga causado por el turismo emergente. Este enfoque permitirá que los resultados y recomendaciones obtenidos puedan ser aplicados en una variedad de contextos, proporcionando beneficios tanto a residentes locales como a visitantes.

## 1.1 Antecedentes

El fenómeno del turismo emergente y sus efectos en la movilidad urbana y rural han sido objeto de atención por parte de las autoridades gubernamentales en diversas partes del mundo. A continuación, se presentan algunos antecedentes relevantes:

- **Crecimiento del Turismo Emergente y su Impacto en la Movilidad:** En los últimos años, el turismo ha experimentado un crecimiento significativo impulsado por la globalización, el desarrollo tecnológico y la mayor accesibilidad a destinos remotos. Este aumento del turismo ha generado una presión considerable sobre las infraestructuras de transporte y la capacidad de gestión de las ciudades y regiones turísticas.
- **La contribución de los Nómadas Digitales:** Un factor importante en este crecimiento es el fenómeno de los nómadas digitales, profesionales que trabajan de forma remota y aprovechan la conectividad digital para viajar y trabajar desde diferentes ubicaciones. Este grupo demográfico ha contribuido notablemente al incremento del turismo, añadiendo una capa adicional de complejidad a la gestión de la movilidad en estos destinos.

## 1.2 Desafíos

El crecimiento del turismo ha planteado desafíos significativos en términos de movilidad urbana y rural, afectando tanto a residentes como a visitantes.

- **Efectos del turismo emergente en la movilidad:** El notable aumento del turismo ha provocado una congestión considerable en las áreas urbanas y rurales, afectando negativamente la movilidad de residentes y visitantes. La escasez de

estacionamiento y el aumento del tráfico han generado problemas de accesibilidad y tiempos de desplazamiento prolongados. Este exceso de turismo ha llevado a una mayor demanda de movilidad flexible y servicios adaptados a sus necesidades específicas.

- **Desarrollo de soluciones innovadoras:** Ante estos desafíos, se han desarrollado diversas iniciativas y proyectos para mejorar la gestión de la movilidad en destinos turísticos emergentes. Estas soluciones van desde la implementación de sistemas de transporte público eficientes hasta la adopción de tecnologías inteligentes, como la monitorización del tráfico en tiempo real y la gestión dinámica del estacionamiento.
- **Plataformas de datos para la gestión de la movilidad:** Las plataformas de datos han emergido como herramientas clave para recopilar, analizar y visualizar información relevante sobre la movilidad urbana y rural. Estas plataformas permiten la integración de datos provenientes de diversas fuentes, como cámaras de tráfico, sistemas de transporte público y aplicaciones móviles, facilitando la toma de decisiones informadas por parte de las autoridades encargadas.

# Capítulo 2 Plataformas de datos

Durante los últimos años, la computación en la nube ha revolucionado la manera en que las organizaciones manejan y procesan la información. La adopción de servicios en la nube ha permitido a empresas de todos los tamaños escalar sus operaciones, reducir costos, y aumentar la agilidad. Las plataformas de computación en la nube ofrecen una amplia gama de servicios que van desde la infraestructura básica hasta avanzadas capacidades de inteligencia artificial y análisis de datos.

Entre los principales proveedores de servicios en la nube se encuentran Microsoft Azure, Amazon Web Services (AWS) y Google Cloud Platform (GCP). Estas plataformas lideran el mercado y son reconocidas por su innovación, seguridad y capacidad de integración con diversas tecnologías y servicios. Cada una de ellas ofrece un conjunto de características y beneficios, lo que las convierte en opciones atractivas para diferentes necesidades empresariales y técnicas.

Uno de los primeros objetivos de este proyecto es llevar a cabo una evaluación de las tres principales plataformas de computación en la nube: Microsoft Azure, Amazon Web Services y Google Cloud Platform. Esta evaluación se centrará en comparar sus características, servicios, ventajas y desventajas con el fin de poder elegir la mejor plataforma que se ajuste a las necesidades e intereses planteados durante el proyecto.

## 2.1 Definición y requisitos

Una plataforma de datos es un conjunto de servicios, herramientas y recursos diseñados para ayudar a las organizaciones a gestionar, almacenar, procesar, analizar y visualizar datos de manera eficiente y segura. Estas plataformas están concebidas para permitir a las empresas maximizar el valor de sus datos, proporcionando información valiosa que facilita la toma de decisiones informadas y estratégicas. Entre los servicios comunes que ofrece una plataforma de datos se incluyen bases de datos relacionales y no relacionales, almacenamiento en la nube, análisis de big data, inteligencia artificial y machine learning, así como herramientas de visualización de datos.

Los requisitos pueden variar dependiendo de las necesidades específicas del usuario y del uso que se le va a dar a los datos. Este trabajo de fin de grado se centra en los requisitos no funcionales ya que el estudio se centra en las características de la plataforma de datos y no en la aplicación que se desarrollará en la misma. Esta aplicación tiene una ingesta de datos, un procesamiento y una visualización a través de cuadros de mando.

En términos generales, los requisitos no funcionales que suelen considerarse al elegir una plataforma de datos incluyen:

- **Escalabilidad:** La plataforma debe ser capaz de escalar horizontal y verticalmente para manejar grandes volúmenes de datos a medida que la empresa crece y evoluciona. Esto incluye la capacidad de ampliar tanto el almacenamiento como el procesamiento de datos sin comprometer el rendimiento.
- **Rendimiento:** Debe ofrecer un rendimiento óptimo para procesar y analizar datos de manera rápida y eficiente. Esto incluye tiempos de respuesta bajos, alta capacidad de procesamiento y la posibilidad de realizar análisis en tiempo real.
- **Seguridad:** Es fundamental que la plataforma garantice la seguridad de los datos, incluyendo cifrado en tránsito y en reposo, gestión de identidades y accesos, y cumplimiento con las normativas de privacidad y protección de datos (como RGPD).
- **Integración:** Debe ser compatible con las herramientas y sistemas existentes del

usuario, permitiendo una fácil integración con otras aplicaciones y sistemas de datos. Esto incluye compatibilidad con las APIs, conectores y herramientas de ETL (extracción, transformación y carga de datos).

- **Disponibilidad y Fiabilidad:** Debe ofrecer una alta disponibilidad y fiabilidad para garantizar la continuidad del negocio y la recuperación ante desastres. Esto se logra mediante mecanismos de redundancia, copias de seguridad automáticas y políticas de recuperación ante fallos.
- **Costo:** El costo de la plataforma, incluyendo licencias, tarifas de suscripción, costos de almacenamiento y procesamiento, debe ser considerado para asegurar que se ajuste al presupuesto disponible. Además, es importante evaluar el costo total de propiedad, que incluye costos operativos y de mantenimiento a largo plazo.
- **Soporte y mantenimiento:** La calidad del soporte técnico y las opciones de mantenimiento disponibles también son factores cruciales. Es importante que la plataforma ofrezca soporte 24/7, documentación exhaustiva y acceso a recursos de capacitación y comunidad.
- **Flexibilidad y personalización:** La capacidad de personalizar y configurar la plataforma según las necesidades específicas del usuario puede ser un factor diferenciador. Esto incluye la posibilidad de crear soluciones a medida y adaptar la plataforma a los cambios requeridos del negocio.

## 2.2 Microsoft Azure

Azure [\[1\]](#) es la plataforma en la nube de Microsoft que ofrece una amplia gama de servicios y soluciones para empresas de todos los tamaños. Está compuesta por más de 200 productos y servicios en la nube, permitiendo trabajar de forma híbrida, ya sea en un entorno local o en varias nubes. Con más de 90 ofertas de cumplimiento normativo, Microsoft invierte 1.000 Millones de dólares estadounidenses al año en seguridad para proteger los datos de los clientes frente a amenazas.

Entre la gran variedad de productos que ofrece Azure, los más destacados son:

- **Azure Open Datasets:** Plataforma en la nube para hospedar y compartir conjuntos de datos abiertos organizados para agilizar el desarrollo de modelos de Machine Learning.
- **Azure data explorer:** Servicio de análisis de datos totalmente administrado para analizar grandes volúmenes de streaming de datos procedentes de aplicaciones, sitios web, dispositivos IoT, etc., en tiempo real.
- **Microsoft Fabric:** Microsoft Fabric es una plataforma de datos y análisis integral diseñada para empresas que requieren una solución unificada. Abarca el movimiento de datos, el procesamiento, la ingesta, la transformación, el enrutamiento de eventos en tiempo real y la creación de informes.
- **Azure Maps:** Agrega datos de ubicación y objetos visuales de asignación a soluciones y aplicaciones.
- **Azure HDInsight:** Servicio de Apache Hadoop administrado que permite ejecutar Apache Spark, Apache Hive, Apache Kafka, Apache HBase y más en la nube.
- **Azure Synapse Analytics:** Servicio de análisis empresarial que acelera el tiempo necesario para obtener información de los sistemas de almacenamiento de datos y macrodatos.

Para Microsoft Azure, no existen paquetes predeterminados para comprar; en cambio, los usuarios/organizaciones pueden elegir y combinar los servicios según sus

necesidades específicas. Los costos se calculan utilizando una calculadora de precios [2] que permite a los usuarios estimar los gastos basados en los servicios seleccionados y el uso previsto. Azure utiliza un modelo de pago por uso, donde los costos están determinados por el consumo de recursos como la computación, el almacenamiento y el ancho de banda. Esto permite a las organizaciones y a los desarrolladores optimizar sus gastos en función de sus necesidades cambiantes y solo pagar por lo que realmente utilizan.

## 2.3 Amazon Web Services

Amazon Web Services (AWS) [3] es una de las nubes más completas del mundo, ofreciendo soluciones escalables para la computación, el almacenamiento, las bases de datos, el análisis y mucho más. AWS ofrece tanto infraestructura como servicio (IaaS) como plataforma como servicio (PaaS).

AWS destaca por su robusta infraestructura y sus amplias opciones de capacitación y certificación para mejorar habilidades técnicas y experiencia en la nube. Está diseñado para ser la infraestructura en la nube global más segura, apoyando 143 estándares de seguridad y certificaciones de conformidad, ayudando a los clientes a cumplir con los requisitos de conformidad en todo el mundo.

Entre la gran variedad de productos que ofrece AWS, los más destacados son:

- **AWS Glue:** Servicio de integración de datos sin servidor que brinda varios motores de integración de datos.
- **Amazon Managed Streaming for Apache Kafka (MSK):** Servicio Apache Kafka de alta disponibilidad y completamente administrado para transmitir datos de forma segura.
- **Amazon Redshift:** Almacenamiento de datos rápido, sencillo y rentable.
- **Amazon API Gateway:** Servicio para crear, desplegar y administrar APIs a cualquier escala

Similar a Microsoft Azure, AWS permite a los usuarios seleccionar y pagar solo por los servicios que necesitan. Los costos se pueden calcular utilizando una calculadora de precios [4], y los usuarios pueden solicitar un presupuesto personalizado contactando con un especialista disponible en la plataforma.

## 2.4 Google Cloud Platform

Google Cloud Platform (GCP) [5] es la suite de infraestructuras y servicios que Google utiliza a nivel interno y está disponible para cualquier empresa. GCP reúne las innovaciones de todos los ámbitos de Google para ayudar a los clientes a llevar a cabo su transformación digital con una IA preparada para el futuro, optimizando la estrategia de IA sin poner en riesgo la confidencialidad de los datos.

GCP ayuda a los desarrolladores a realizar su trabajo de forma rápida, segura y rentable con una infraestructura moderna de nueva generación, diseñada para adaptarse a cargas de trabajo específicas y a las necesidades de cada sector. Además, GCP cumple con los principales estándares normativos.

Entre los más de 100 productos que ofrece Google Cloud Platform, los más destacados son:

- **BigQuery:** Almacén de datos sin servidores que funciona en varias nubes y escala con tus datos, utilizando IA y análisis avanzados para obtener estadísticas a gran escala.
- **Vertex AI:** Plataforma unificada para modelos de aprendizaje automático e IA

generativa.

- **Looker:** Plataforma de BI, aplicaciones de datos y analíticas integradas.
- **Cloud SQL:** Servicio de bases de datos relacionales para MySQL, PostgreSQL y SQL Server.
- **Google Maps Platform:** Servicio para mostrar o analizar datos geoespaciales en un mapa para facilitar la toma de decisiones.
- **Dataproc:** Servicio totalmente gestionado y altamente escalable para ejecutar Apache Hadoop y más de 30 herramientas y frameworks de software libre, permitiendo integrarse con Vertex AI, BigQuery y Dataplex.

Al igual que en las plataformas anteriores, en Google Cloud Platform solo se paga por los servicios necesarios. Proporciona una calculadora de precios para realizar estimaciones mensuales del costo de los servicios utilizados [6]. Además, los usuarios pueden contactar a un especialista para obtener una solución personalizada basada en sus necesidades específicas.

## 2.5 Conclusiones de las plataformas de datos

Tras haber evaluado las principales plataformas en la nube disponibles, Microsoft Azure se destaca como la opción más adecuada para nuestro proyecto. Basándonos en los servicios proporcionados por Azure y las necesidades e intereses específicos de nuestro proyecto, hemos decidido centrar nuestros esfuerzos en esta plataforma por varias razones.

En primer lugar, la extensa gama de servicios que ofrece Azure. Con más de 200 productos y servicios en la nube, Azure proporciona soluciones integrales que abarcan desde el almacenamiento y procesamiento de datos hasta la inteligencia artificial y el aprendizaje automático. Esta amplia gama de servicios nos permite tener la flexibilidad de seleccionar y combinar servicios que se alinean perfectamente con los requisitos de nuestro proyecto.

Microsoft Fabric se destaca como una herramienta esencial para nuestro proyecto. Esta plataforma de datos nos permite obtener, crear, compartir y visualizar datos de manera eficaz. Fabric facilita la integración de diversos datos y la creación de soluciones personalizadas, lo cual es crucial para nuestra capacidad de innovar y adaptarnos a las necesidades cambiantes del proyecto.

La estructura de costos de Azure también juega un papel importante en nuestra decisión. Al no existir paquetes predeterminados, podemos elegir los servicios que realmente necesitamos y ajustar nuestros gastos en función de nuestras necesidades específicas. La calculadora de precios en línea de Azure nos permite planificar financieramente, proporcionando estimaciones detalladas de los costos basados en nuestras configuraciones de servicios. El modelo de pago por uso de Azure asegura que solo pagamos por los recursos que consumimos, optimizando así nuestro presupuesto y evitando costos innecesarios.

En resumen, Microsoft Azure se adecua perfectamente con nuestras necesidades técnicas. La plataforma no solo nos proporciona los servicios y herramientas necesarios para llevar a cabo nuestro proyecto de manera efectiva, sino que también nos ofrece la seguridad, flexibilidad y soporte que necesitamos para garantizar el éxito a largo plazo. Con la inclusión de Microsoft Fabric como una herramienta central para la gestión y visualización de datos, estamos mejor posicionados para innovar y alcanzar nuestros objetivos. Por estas razones, se ha decidido que Microsoft Azure es la plataforma ideal para nuestro proyecto y esperamos aprovechar al máximo las oportunidades que ofrece.

# Capítulo 3 Microsoft Fabric

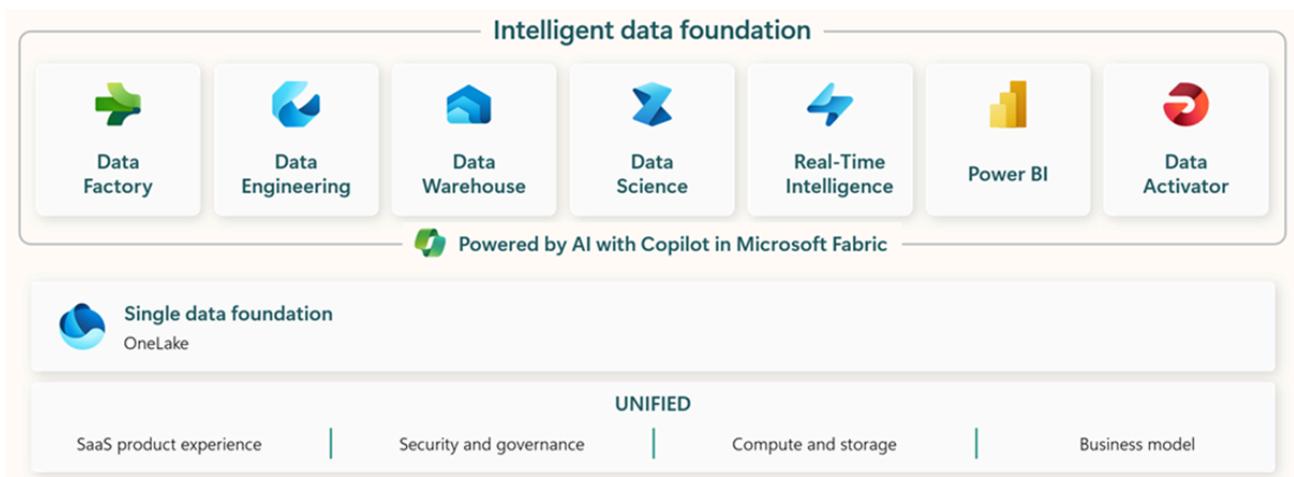
Tras haber evaluado las distintas plataformas en la nube y los servicios que ofrece, decidimos centrarnos en Microsoft Azure, concretamente en Microsoft Fabric para el desarrollo de nuestro proyecto. En este capítulo detallaremos cómo Fabric es una plataforma de datos que integra diversos servicios de Azure en un entorno unificado.

## 3.1 Definición

Microsoft Fabric es una plataforma de datos y análisis integral diseñada para empresas que requieren una solución unificada. Abarca el movimiento de datos, el procesamiento, la ingesta, la transformación, el enrutamiento de eventos en tiempo real y la creación de informes. Ofrece un conjunto de servicios, como Data Engineering, Data Factory, Ciencia de datos, Real-Time Analytics, Data Warehouse y Bases de datos.

Con Fabric, no es necesario utilizar diferentes servicios de varios proveedores. En su lugar, contará con una plataforma perfectamente integrada y fácil de usar. Al funcionar con un modelo de software como servicio (SaaS), Fabric aporta simplicidad e integración a sus soluciones.

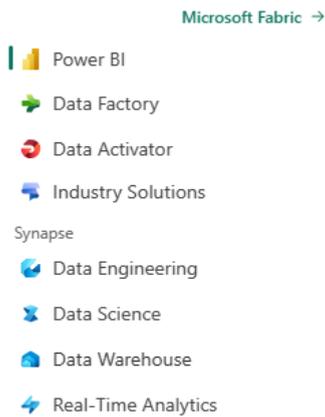
Microsoft Fabric integra componentes independientes en una pila cohesiva. En lugar de depender de bases de datos o almacenes de datos diferentes, puede centralizar el almacenamiento de datos con OneLake. Las funcionalidades de inteligencia artificial se insertan sin problemas en Fabric, lo que elimina la necesidad de realizar una integración manual. Con Fabric, podemos realizar fácilmente la transición de los datos sin procesar a información procesable para usuarios empresariales.



**Figura 3.1:** *Arquitectura de Microsoft Fabric.*

Fabric integra cargas de trabajo como Ingeniería de datos, Data Factory, Ciencia de datos, Data Warehouse, Inteligencia en tiempo real, Industry Solutions y Power BI en una base de SaaS compartida. Cada una de estas experiencias se adapta a roles de usuario distintos, como ingenieros de datos, científicos o profesionales de almacenamiento, y sirven para una tarea específica. Toda la pila de Fabric tiene integración con IA y acelera el recorrido de los datos.

Fabric ofrece un conjunto completo de experiencias de análisis diseñadas para trabajar conjuntamente sin problemas. La plataforma adapta cada una de estas experiencias a un rol específico y a una tarea específica que explicaremos a continuación.



**Figura 3.2:** *Componentes de Microsoft Fabric.*

## 3.2 Power BI

Power BI es una colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente. Sus datos podrían ser una hoja de cálculo de Excel o una colección de almacenes de datos híbridos locales y basados en la nube. Power BI permite conectarse con facilidad a los orígenes de datos, visualizar y descubrir qué es importante y compartirlo con cualquiera o con todos los usuarios que desee.

Se conforma fundamentalmente de estos componentes:

- Power BI Desktop: aplicación gratuita de escritorio para transformar, visualizar datos y crear informes de los mismos.
- Power BI Service: servicio online (SaaS) con funcionalidad similar a la aplicación desktop y permite publicar informes y configurar la actualización de datos automáticamente para que el personal de la organización tenga los datos actualizados.
- Power BI Mobile: aplicación móvil disponible para Windows, iOS y Android para visualizar informes y que se actualiza automáticamente con los cambios de los datos.

Power BI permite conectar a cientos de orígenes de datos en la nube o entorno local, creando informes con objetos integrados o creando objetos personalizados.

El acceso a los datos puede ser desde Microsoft Fabric, una tabla Excel, Salesforce, Dynamic CRM, Google Analytics, hasta complejas bases de datos (on-premise o en la nube), información de servicios de Azure, etc., lo cual facilita tener toda la información en una única visualización.

Con el análisis de los datos, se pueden evaluar los crecimientos, debilidades, fortalezas de una organización, y permitir la toma de acertadas decisiones a corto, mediano y largo plazo, a través de los paneles, informes y gráficos interactivos.

Power BI, nos permite analizar los datos y obtener patrones “poco visibles” y que ayuden a llegar a conclusiones y toma de decisiones en favor de nuestra organización. Para ello se puede usar la previsión, la agrupación, las medidas rápidas; y para usuarios más avanzados esta herramienta brinda un potente lenguaje de fórmulas de DAX, que nos proporciona un control total sobre el modelo.

Una de las mayores ventajas de esta herramienta es su integración con Microsoft Fabric, pues los datos se pueden exportar y conectar con los paneles de Power BI fácilmente y como Power BI se actualiza constantemente, el análisis de los datos es generado en tiempo real, permitiendo la toma de decisiones al momento.

Podemos decir que Power BI es una buena herramienta para los análisis a nivel empresarial. Su flexibilidad para extraer información de una gama amplia de fuentes y ser utilizado por diferentes sectores: económicos, salud, industria, informática y comunicaciones, etc.; es innovador porque permite el desarrollo de informes creativos y la visualización interactiva de los datos; al ser multiplataforma permite crear informes para dispositivos móviles; además combinando, optimizando y transformando datos su optimización le permite un análisis profundo de estos encontrando patrones en los mismos. Además, como los datos están securizados no habrá problemas si se acceden desde un dispositivo móvil, pues el proceso de autenticación del usuario se realiza a través de Azure Active Directory.

### 3.3 Data Factory

Data Factory permite disfrutar de una experiencia de integración de datos moderna para ingerir, preparar y transformar datos de un amplio conjunto de orígenes de datos (por ejemplo, bases de datos, almacenamiento de datos, *Lakehouse*, datos en tiempo real, etc.). Para cualquier tipo de rol, se pueden transformar los datos con transformaciones inteligentes y aprovechar un amplio conjunto de actividades.

Podemos crear canalizaciones para ejecutar una o varias actividades, acceder a orígenes de datos o servicios a través de servicios vinculados y, después de crear una canalización, podemos agregar desencadenadores para ejecutar automáticamente nuestros procesos en momentos específicos o en respuesta a escenarios cambiantes. Con Data Factory en Microsoft Fabric, incorporamos funcionalidades de copia rápida (movimiento de datos) a flujos de datos y canalizaciones de datos.

Hay dos características principales de alto nivel que Data Factory implementa: flujos de datos y canalizaciones.

- **Flujos de datos:** Los flujos de datos permiten aprovechar más de 300 transformaciones en el diseñador de flujos de datos, lo que le permite transformar los datos más fácilmente y con más flexibilidad que cualquier otra herramienta, incluidas las transformaciones de datos basadas en inteligencia artificial inteligente.
- **Canalizaciones de datos:** Las canalizaciones de datos permiten aprovechar las funcionalidades de orquestación de datos enriquecidas de fábrica para crear flujos de trabajo de datos flexibles que satisfagan sus necesidades empresariales.

Puesto que Microsoft Fabric se encuentra en constante evaluación, actualmente ofrece características emergentes que están en fase de previsualización. Entre ellas se encuentran el flujo de trabajo de datos con Apache Airflow y la API para GraphQL.

- **Flujo de trabajo de datos con Apache Airflow:** Esta funcionalidad permite la integración de Apache Airflow, una herramienta de orquestación de flujos de trabajo de código abierto, para crear flujos de trabajo complejos y personalizados. Aunque esta característica aún está en fase de previsualización y no está completamente operativa, ofrece un vistazo prometedor a las capacidades avanzadas de orquestación que podemos aprovechar en el futuro.
- **API para GraphQL:** GraphQL es un lenguaje de consulta y un tiempo de ejecución del servidor para las interfaces de programación de aplicaciones (API). Con GraphQL, las API son rápidas, flexibles y sencillas para los desarrolladores. Incluso se puede implementar en un entorno de desarrollo integrado (IDE) conocido como GraphiQL. Como alternativa a REST, GraphQL permite que los desarrolladores creen consultas para extraer datos de varias fuentes en una sola llamada a la API. Aunque la API para GraphQL en Microsoft Fabric está en fase de previsualización y no está totalmente operativa, su potencial para mejorar nuestras capacidades de consulta y manipulación de datos es considerable.

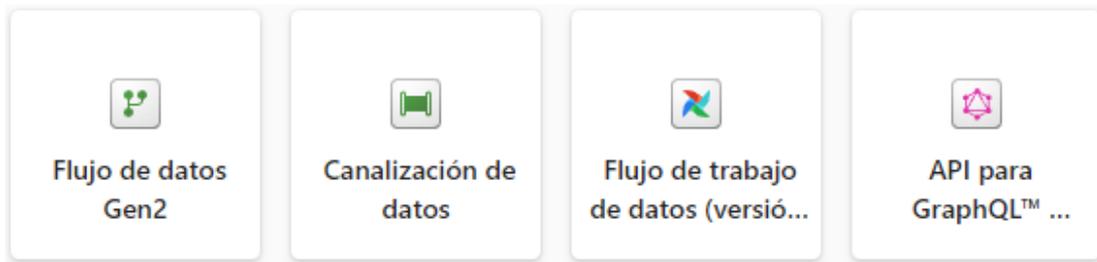


Figura 3.3: Funcionalidades de Data Factory.

### 3.3.1 Flujo de datos Gen2

Los flujos de datos proporcionan una interfaz de bajo código para ingerir datos de cientos de orígenes de datos, lo que transforma los datos con más de 300 tipos de transformaciones de datos. A continuación, puede cargar los datos resultantes en varios destinos, como Azure SQL bases de datos, datalakes y mucho más. Los flujos de datos se pueden ejecutar repetidamente mediante la actualización manual o programada, o como parte de una orquestación de canalización de datos.

Los flujos de datos se crean utilizando la experiencia familiar de Power Query que está disponible hoy en día en varios productos y servicios de Microsoft, como Excel, Power BI, Power Platform, aplicaciones Dynamics 365 Insights y más. Power Query permite a todos los usuarios realizar ingesta de datos y transformaciones de datos en todo su patrimonio de datos. Podemos hacer combinaciones, agregaciones, limpieza de datos, transformaciones personalizadas y mucho más desde una interfaz de usuario de código bajo, muy visual y fácil de usar.

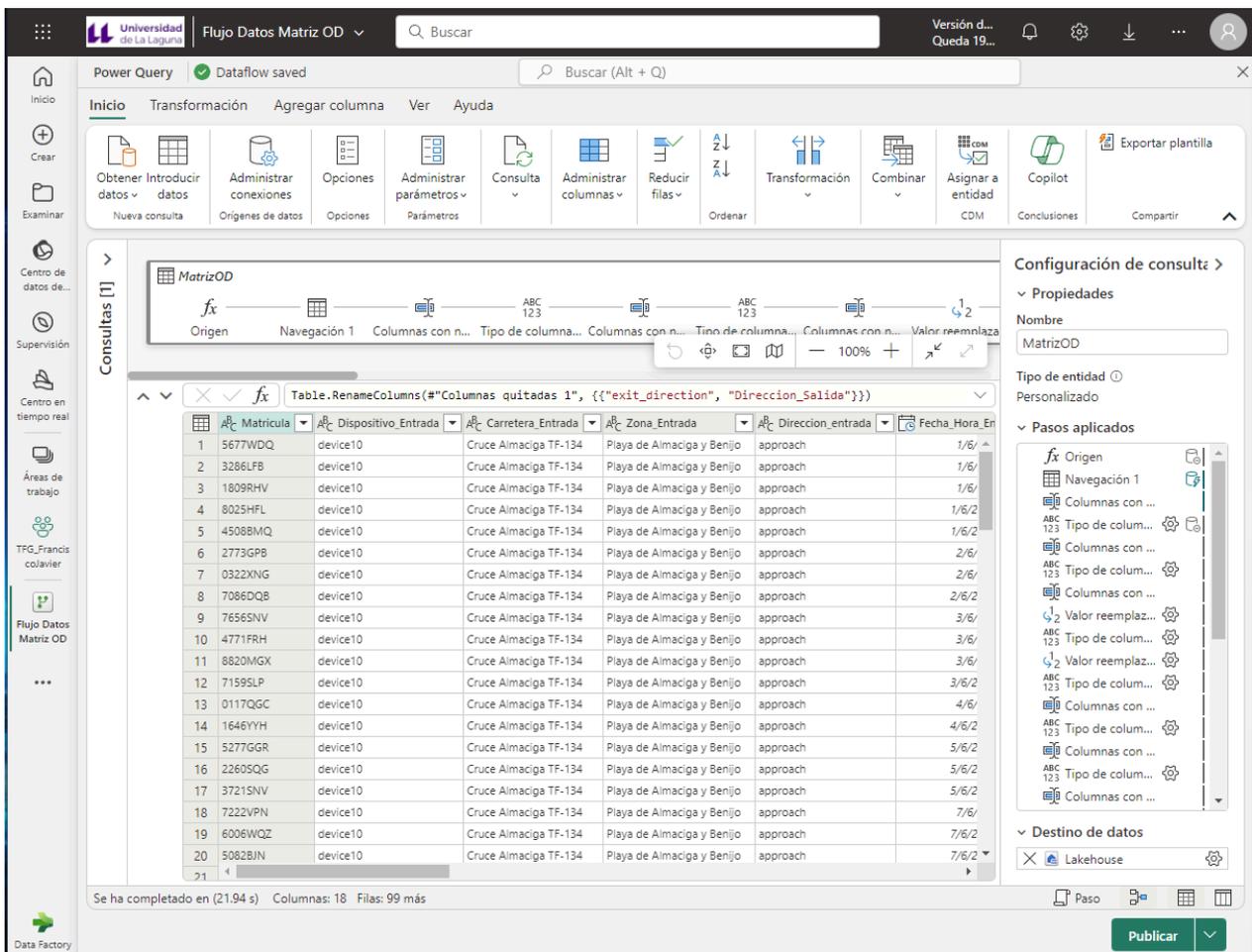
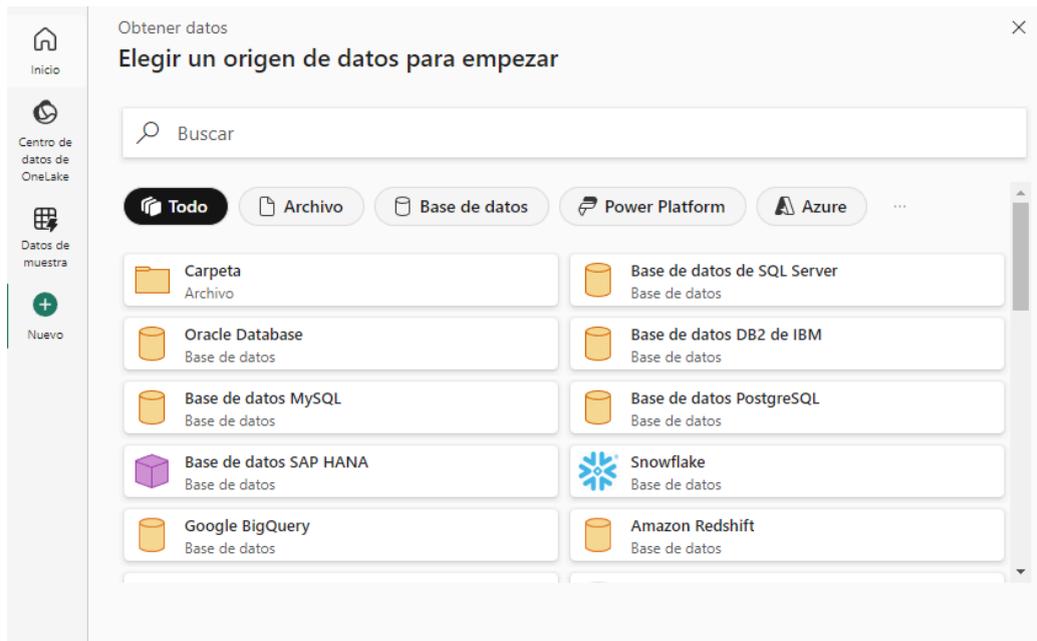


Figura 3.4: Ejemplo de un flujo de datos.

### 3.3.2 Canalizaciones de datos

Las canalizaciones de datos permiten eficaces funcionalidades de flujo de trabajo a escala en la nube. Con las canalizaciones de datos, puede crear flujos de trabajo complejos que puedan actualizar el flujo de datos, mover datos de tamaño PB y definir canalizaciones de flujo de control sofisticadas.

Podemos usar las canalizaciones de datos para crear flujos de trabajo ETL y factoría de datos complejos que pueden realizar muchas tareas diferentes a escala. Las funcionalidades de flujo de control están integradas en canalizaciones de datos que permiten crear lógica de flujo de trabajo, que proporciona bucles y condicionales.



**Figura 3.5:** Pantalla de opción para obtener orígenes de datos.

Agregar una actividad de copia controlada por la configuración junto con la actualización del flujo de datos de poco código en una sola canalización para una canalización de datos ETL de un extremo a otro. Incluso puede agregar actividades de código primero para cuadernos de Spark, scripts SQL, procedimientos almacenados, etc.

Id. de ejecución de canalización: 5b14e958-0136-4c56-b7d7-5204622e5526

Estado de la canalización: ✔ Correcto

Mostrando elementos del 1 al 4

| Nombre de actividad            | Estado de actividad                           | Inicio de la ejecución | Duración | Entrada | Salida |
|--------------------------------|---|------------------------|----------|---------|--------|
| Actualización del modelo se... | <span style="color: green;">✔</span> Correcto | 7/1/2024, 4:22:40 PM   | 19s      | →       | ↳      |
| Flujo datos Matriz OD          | <span style="color: green;">✔</span> Correcto | 7/1/2024, 4:21:06 PM   | 1m 33s   | →       | ↳      |
| Matriz Origen-Destino          | <span style="color: green;">✔</span> Correcto | 7/1/2024, 4:20:40 PM   | 26s      | →       | ↳      |
| Simulador                      | <span style="color: green;">✔</span> Correcto | 7/1/2024, 4:20:06 PM   | 33s      | →       | ↳      |

Figura 3.6: Ejemplo de canalizador de datos

## 3.4 Data Activator

Aunque aún está en una versión preliminar, data activator es una experiencia sin código en Microsoft Fabric para realizar acciones automáticamente cuando se detectan patrones o condiciones en los datos cambiantes. Supervisa los datos de los informes de Power BI y los elementos Eventstreams, para cuando los datos alcanzan determinados umbrales o coinciden con otros patrones. A continuación, realiza automáticamente las acciones adecuadas, como alertar a los usuarios o iniciar flujos de trabajo de Power Automate.

El activador de datos permite a los clientes crear un sistema nervioso digital que actúa en todos sus datos, a escala y de forma oportuna. Los usuarios empresariales pueden describir las condiciones empresariales en una experiencia sin código para iniciar acciones como correo electrónico, notificaciones de Teams, flujos de Power Automate y llamar a sistemas de acción de terceros.

### 3.4.1 Eventos

Data Activator considera que todos los orígenes de datos son flujos de eventos. Un evento es una observación sobre el estado de un objeto, con algún identificador para el propio objeto, una marca de tiempo y los valores de los campos que está supervisando. Los flujos de eventos varían en frecuencia desde muchas veces por segundo en el caso de los sensores de IoT hasta secuencias más esporádicas, como los paquetes que se escanean al salir y al entrar en las ubicaciones de envío.

Los datos que se observan desde Power BI también se tratan como un flujo de eventos. En este caso, los eventos son observaciones que se hacen de los datos en una programación normal contra el modelo semántico de Power BI (anteriormente conocido como conjunto de datos). Esta consulta se produce una vez por hora: es simplemente un flujo de eventos que cambia lentamente.

### 3.4.2 Objetos

Los objetos representan la definición o clase de una entidad que se desea supervisar (vehículos, usuarios...). El término "instancia de objeto" se refiere a un elemento específico de una entidad, como un vehículo concreto, usuario determinado, mientras que "objeto" se utiliza para definir la clase o el tipo del objeto en sí. El concepto de "población"

se refiere a la totalidad de las instancias de un objeto. Este enfoque permite una gestión granular y precisa de los datos empresariales, facilitando el seguimiento y análisis de cada instancia y su comportamiento a lo largo del tiempo.

### **3.4.3 Desencadenadores**

Los desencadenadores definen las condiciones que desea detectar en los objetos y las acciones que desea realizar cuando se cumplen esas condiciones. Cuando se activa un desencadenador, siempre es para una instancia de objeto específica. Un desencadenador en un objeto congelador podría detectar que el congelador está demasiado cálido y enviar un correo electrónico al técnico correspondiente.

### **3.4.4 Propiedades**

Las propiedades son útiles cuando se desea reutilizar la lógica en varios desencadenadores. Puede definir una propiedad en un objeto congelador que suaviza las lecturas de temperatura durante un período de una hora. Después, podría usar ese valor suavizado en muchos desencadenadores.

## **3.5 Industry solutions**

Microsoft Fabric ofrece soluciones de datos específicas de sector que proporcionan una plataforma sólida para la administración de datos, el análisis y la toma de decisiones. Estas soluciones de datos abordan los desafíos únicos a los que se enfrentan diferentes sectores, lo cual permite a las empresas optimizar las operaciones, integrar datos de diferentes orígenes y utilizar análisis enriquecidos.

Una de las características clave de Microsoft Fabric Industry Solutions es la personalización de soluciones según la industria. Por ejemplo, en el sector salud, las soluciones incluyen herramientas para la gestión de datos de pacientes, análisis de resultados clínicos y optimización de recursos hospitalarios. Para la industria financiera, se proporcionan soluciones para el análisis de riesgos, cumplimiento regulatorio y personalización de servicios financieros mediante el uso de datos de clientes y transacciones.

La integración de datos es otro aspecto crucial de Microsoft Fabric Industry Solutions. La plataforma permite la integración de datos de diversas fuentes, incluyendo sistemas de gestión empresarial (ERP), sistemas de gestión de relaciones con clientes (CRM), dispositivos IoT y plataformas de comercio electrónico. Esto proporciona una visión holística de la operación empresarial, facilitando la consolidación y limpieza de datos para realizar un análisis más preciso y efectivo. Esta integración asegura que las decisiones se basen en datos completos y actualizados.

La seguridad y el cumplimiento son prioridades esenciales en Microsoft Fabric Industry Solutions. La plataforma asegura que los datos estén protegidos y cumplan con las normativas y estándares de la industria. Proporciona mecanismos robustos de control de acceso y auditoría para garantizar la seguridad y la privacidad de los datos. Esto es particularmente importante en industrias reguladas como la salud y las finanzas, donde el manejo adecuado de datos sensibles es crítico.

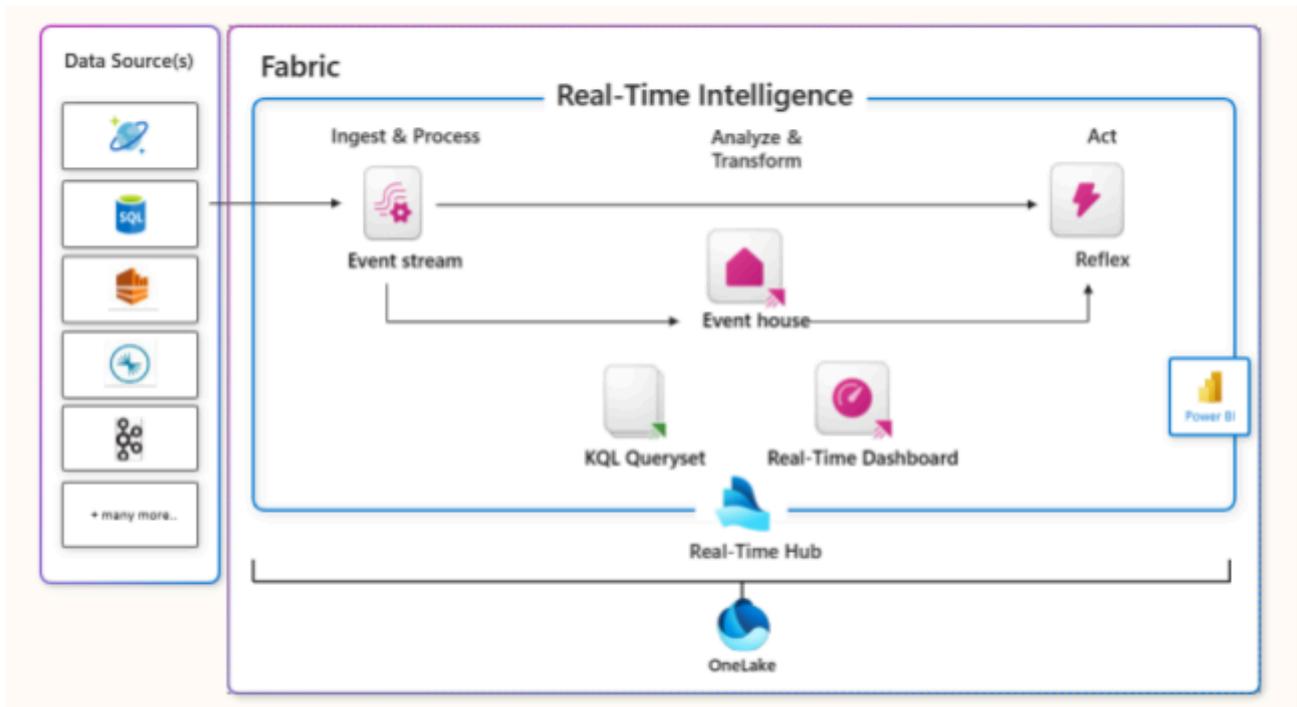
## **3.6 Real-Time Intelligence**

La inteligencia en tiempo real es un servicio eficaz que permite a todos los usuarios de tu organización extraer información y visualizar sus datos en movimiento. Además, ofrece una solución integral para escenarios controlados por eventos, datos de streaming y registros de datos. Tanto si se trata de gigabytes como petabytes, todos los datos de la organización en movimiento convergen en el centro en tiempo real. Conecta sin problemas los datos basados en el tiempo de varios orígenes mediante conectores sin código, lo que permite información visual inmediata, análisis geoespacial y reacciones

basadas en desencadenadores que forman parte de un catálogo de datos de toda la organización.

Una vez nos conectemos a cualquier flujo de datos, se puede acceder a toda la solución SaaS. La inteligencia en tiempo real controla la ingesta de datos, la transformación, el almacenamiento, el análisis, la visualización, el seguimiento, la IA y las acciones en tiempo real. Los datos permanecen protegidos, regulados e integrados en toda la organización, y se alinean sin problemas con todas las ofertas de Fabric. La inteligencia en tiempo real transforma los datos en un recurso dinámico y accionable que impulsa el valor en toda la organización.

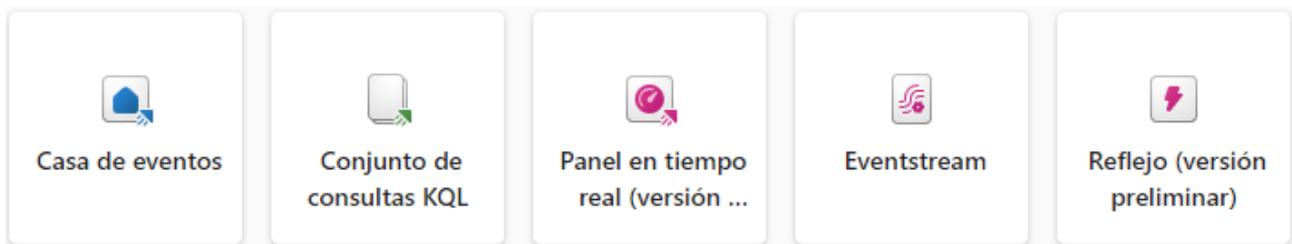
La inteligencia en tiempo real de Microsoft Fabric ofrece funcionalidades que, en combinación, permiten la creación de soluciones de inteligencia en tiempo real en apoyo de procesos empresariales y de ingeniería.



**Figura 3.7:** Diagrama de la arquitectura de inteligencia en tiempo real en Microsoft Fabric.

El centro en tiempo real actúa como un catálogo centralizado dentro de la organización. Facilita el acceso, la adición, la exploración y el uso compartido de datos. Al expandir el intervalo de orígenes de datos, permite información más amplia y claridad visual en varios dominios. Cabe destacar que este centro garantiza que los datos no solo estén disponibles, sino que también sean accesibles para todos, lo que promueve la toma de decisiones rápida y la acción informada. El uso compartido de datos de streaming de diversos orígenes desbloquea el potencial de crear una inteligencia empresarial completa en toda la organización.

Una vez seleccionada una secuencia de la organización o conectada a orígenes externos o internos, puedes usar las herramientas de consumo de datos en Inteligencia en tiempo real para explorar los datos. Las herramientas de consumo de datos usan el proceso de exploración de datos visuales y exploran en profundidad la información de datos. Puedes acceder a los datos que son nuevos y comprender fácilmente la estructura de datos, los patrones, las anomalías, las cantidades de previsión y las tasas de datos. En consecuencia, podemos actuar o tomar decisiones inteligentes en función de los datos. Los paneles en tiempo real vienen equipados con interacciones integradas que simplifican el proceso de comprensión de los datos, lo que hace que sea accesible para cualquier persona que quiera tomar decisiones en función de los datos en movimiento mediante herramientas visuales, lenguaje natural y Copilot.



**Figura 3.8:** *Funcionalidades de Real-Time Intelligence.*

### 3.6.1 Detección de datos de streaming

El centro en tiempo real se usa para detectar y administrar los datos de streaming. Los eventos del centro en tiempo real son un catálogo de datos en movimiento y contienen:

- **Flujos de datos:** se podrá visualizar todas las secuencias de datos que se ejecutan activamente en Fabric a las que tenemos acceso.
- **Orígenes de Microsoft:** es capaz de detectar fácilmente los orígenes de streaming que tenemos y configura rápidamente la ingesta de esos orígenes en Fabric, por ejemplo: Azure Event Hubs, Azure IoT Hub, Azure SQL DB captura de datos modificados (CDC), CDC de Azure Cosmos DB, CDC de PostgreSQL DB.
- **Eventos de Fabric:** las funcionalidades controladas por eventos admiten notificaciones en tiempo real y procesamiento de datos. Podemos supervisar y reaccionar a eventos, incluidos los eventos de elementos del área de trabajo de Fabric y los eventos de Azure Blob Storage. Estos eventos se pueden usar para desencadenar otras acciones o flujos de trabajo, como invocar una canalización de datos o enviar una notificación por correo electrónico. También podemos enviar estos eventos a otros destinos mediante secuencias de eventos.

### 3.6.2 Conexión a datos de streaming

Las secuencias de eventos son la forma en que Fabric captura, transforma y enruta grandes volúmenes de eventos en tiempo real a varios destinos con una experiencia sin código. Los flujos de eventos admiten varios orígenes de datos y destinos de datos, incluida una amplia gama de conectores a orígenes externos, por ejemplo: clústeres de Apache Kafka, fuentes de captura de datos modificados de base de datos, orígenes de streaming de AWS (Kinesis) y Google (GCP Pub/Sub).

### 3.6.3 Procesar flujos de datos

Mediante el uso de las funcionalidades de procesamiento de eventos en flujos de eventos, podemos realizar el filtrado, la limpieza de datos, la transformación, las agregaciones en ventanas y la detección de réplicas para colocar los datos en la forma que deseamos. También podemos usar las funcionalidades de enrutamiento basadas en contenido para enviar datos a diferentes destinos en función de los filtros. Otra característica, derivada de secuencias de eventos, permite construir flujos nuevos como resultado de transformaciones o agregaciones que se pueden compartir con los consumidores en el centro en tiempo real.

### 3.6.4 Almacenar y analizar datos

Los Eventhouse son el motor de análisis ideal para procesar datos en movimiento. Se adaptan específicamente a eventos de streaming basados en tiempo con datos estructurados, semiestructurados y no estructurados. Los datos se indexan y particionan automáticamente en función del tiempo de ingesta, lo que proporciona funcionalidades de consulta analíticas increíblemente rápidas y complejas en datos de granularidad alta. Los datos almacenados en los Eventhouse pueden estar disponibles en OneLake para su consumo por otras experiencias de Fabric.

Los datos indexados y particionados almacenados en los Eventhouse están listos para consultas rápidas con varios códigos, poco código u opciones sin código en Fabric. Los

datos se pueden consultar en KQL nativo (Lenguaje de consulta Kusto) o mediante T-SQL en el conjunto de consultas KQL. El copiloto Kusto, junto con la experiencia de exploración de consultas sin código, simplifica el proceso de análisis de datos para usuarios de KQL experimentados y científicos de datos. KQL es un lenguaje sencillo pero eficaz para consultar datos estructurados, semiestructurados y no estructurados. El lenguaje es expresivo, fácil de leer y comprender la intención de consulta y está optimizado para experiencias de creación.

### 3.6.5 Visualización de información de datos

Estas conclusiones de datos se pueden visualizar en conjuntos de consultas KQL, paneles en tiempo real e informes de Power BI, con segundos de ingesta de datos a información. Las opciones de visualización abarcan desde experiencias sin código hasta experiencias totalmente especializadas. Podemos usar indicaciones visuales para realizar operaciones de filtrado y agregación en los resultados de la consulta y usar una lista enriquecida de visualizaciones integradas. Estas conclusiones se pueden ver en informes de Power BI y paneles en tiempo real, que pueden tener alertas basadas en la información de datos.

### 3.6.6 Acciones de desencadenador

Las alertas supervisan el cambio de datos y realizan acciones automáticamente cuando se detectan patrones o condiciones. Los datos pueden fluir en el centro en tiempo real u observarse desde una consulta de Kusto o un informe de Power BI. Cuando se cumplen determinadas condiciones o lógicas, se realiza una acción, como alertar a los usuarios, ejecutar elementos de trabajo de Fabric como una canalización o iniciar flujos de trabajo de Power Automate. La lógica puede ser un umbral simplemente definido, un patrón como eventos que se producen repetidamente durante un período de tiempo o los resultados de una lógica compleja definida por una consulta KQL.

## 3.7 Data Engineering

Ingeniería de datos en Microsoft Fabric permite a los usuarios diseñar, compilar y mantener infraestructuras y sistemas que permiten a sus organizaciones recopilar, almacenar, procesar y analizar grandes volúmenes de datos.

Microsoft Fabric proporciona diversas funcionalidades de ingeniería de datos para asegurarse de que los datos sean fácilmente accesibles, de alta calidad y estén bien organizados. Desde la página principal de Ingeniería de datos, puede:

- Crear y administrar los datos mediante un data lake.
- Diseñar canalizaciones para copiar datos en el data lake.
- Usar definiciones de trabajos de Spark para enviar un trabajo por lotes o streaming a un clúster de Spark.
- Usar cuadernos para escribir código para la ingesta, preparación y transformación de datos.



Figura 3.9: Funcionalidades de Data Engineering.

### 3.7.1 Lakehouse

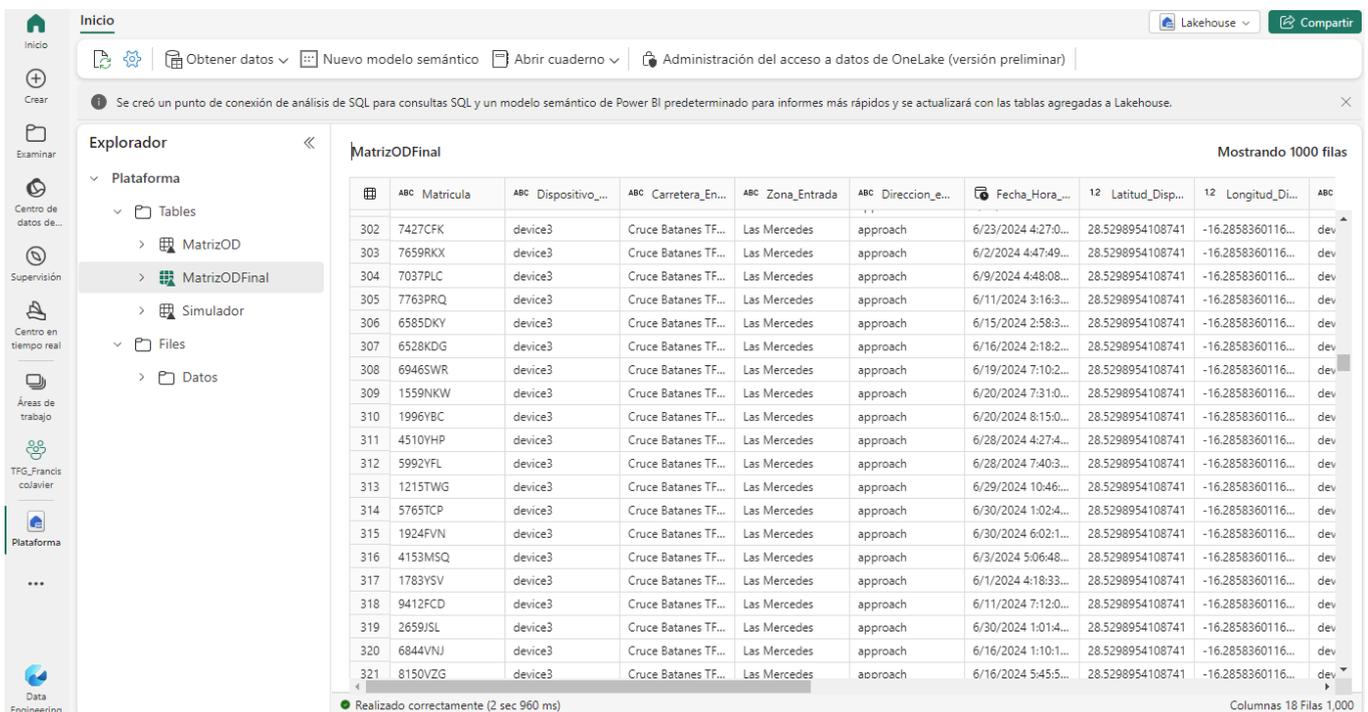
Los almacenes de lago (*Lakehouse*) son arquitecturas de datos que permiten a las organizaciones almacenar y administrar datos estructurados y no estructurados en una sola ubicación, mediante diversas herramientas y marcos para procesar y analizar esos datos. Estas herramientas y marcos pueden incluir consultas y análisis basados en SQL, así como aprendizaje automático y otras técnicas de análisis avanzadas.

El *Lakehouse* crea una capa de servicio generando automáticamente un punto de conexión análisis SQL y un modelo semántico predeterminado durante la creación. Esta nueva funcionalidad transparente permite al usuario trabajar directamente sobre las tablas Delta del lago para ofrecer una experiencia sin fricciones y de alto rendimiento, desde la ingesta de datos hasta la elaboración de informes.

La detección y el registro automáticos de tablas es una característica de *Lakehouse* que proporciona una experiencia de archivo a tabla totalmente administrada para ingenieros de datos y científicos de datos. Puedes soltar un archivo en el área gestionada de *Lakehouse* y el sistema lo valida automáticamente para los formatos estructurados admitidos, y lo registra en el metastore con los metadatos necesarios, como nombres de columnas, formatos, compresión, etc. (Actualmente, el único formato admitido es la tabla Delta). Luego, puede hacer referencia al archivo como una tabla y usar la sintaxis de SparkSQL para interactuar con los datos.

Los ingenieros de datos pueden interactuar con el almacén de lago y los datos dentro de él de distintas maneras:

- **El explorador de Lakehouse:** el explorador es la página de interacción principal de *Lakehouse*. Con el explorador de objetos, puedes cargar datos y explorarlos en *Lakehouse*, establecer etiquetas MIP y mucho más.



The screenshot shows the Lakehouse Explorer interface. On the left is a navigation pane with 'Inicio', 'Crear', 'Examinar', 'Centro de datos de...', 'Supervisión', 'Centro en tiempo real', 'Áreas de trabajo', 'TFG\_Francis colavier', and 'Plataforma'. The main area displays a table named 'MatrizODFinal' with 18 columns and 1,000 rows. The columns are: ABC, Matricula, ABC Dispositivo..., ABC Carretera\_En..., ABC Zona\_Entrada, ABC Direccion\_e..., Fecha\_Hora..., 12 Latitud\_Dis..., 12 Longitud\_Di..., and ABC. The table contains data for various devices and locations, such as 'Cruce Batanes TF...' and 'Las Mercedes'. A status bar at the bottom indicates 'Realizado correctamente (2 sec 960 ms)' and 'Columnas 18 Filas 1,000'.

| ABC | Matricula | ABC Dispositivo... | ABC Carretera_En... | ABC Zona_Entrada | ABC Direccion_e... | Fecha_Hora...       | 12 Latitud_Disp... | 12 Longitud_Di... | ABC |
|-----|-----------|--------------------|---------------------|------------------|--------------------|---------------------|--------------------|-------------------|-----|
| 302 | 7427CFK   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/23/2024 4:27:0... | 28.5298954108741   | -16.2858360116... | dev |
| 303 | 7659RKK   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/2/2024 4:47:49... | 28.5298954108741   | -16.2858360116... | dev |
| 304 | 7037PLC   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/9/2024 4:48:08... | 28.5298954108741   | -16.2858360116... | dev |
| 305 | 7763PRQ   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/11/2024 3:16:3... | 28.5298954108741   | -16.2858360116... | dev |
| 306 | 6585DKY   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/15/2024 2:58:3... | 28.5298954108741   | -16.2858360116... | dev |
| 307 | 6528KDG   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/16/2024 2:18:2... | 28.5298954108741   | -16.2858360116... | dev |
| 308 | 6946SWR   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/19/2024 7:10:2... | 28.5298954108741   | -16.2858360116... | dev |
| 309 | 1559NKW   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/20/2024 7:31:0... | 28.5298954108741   | -16.2858360116... | dev |
| 310 | 1996VBC   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/20/2024 8:15:0... | 28.5298954108741   | -16.2858360116... | dev |
| 311 | 4510VHP   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/28/2024 4:27:4... | 28.5298954108741   | -16.2858360116... | dev |
| 312 | 5992VFL   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/28/2024 7:40:3... | 28.5298954108741   | -16.2858360116... | dev |
| 313 | 1215TWG   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/29/2024 10:46...  | 28.5298954108741   | -16.2858360116... | dev |
| 314 | 5765TCP   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/30/2024 1:02:4... | 28.5298954108741   | -16.2858360116... | dev |
| 315 | 1924FVN   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/30/2024 6:02:1... | 28.5298954108741   | -16.2858360116... | dev |
| 316 | 4153MSQ   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/3/2024 5:06:48... | 28.5298954108741   | -16.2858360116... | dev |
| 317 | 1783YSV   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/1/2024 4:18:33... | 28.5298954108741   | -16.2858360116... | dev |
| 318 | 9412FCD   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/11/2024 7:12:0... | 28.5298954108741   | -16.2858360116... | dev |
| 319 | 2659JSL   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/30/2024 1:01:4... | 28.5298954108741   | -16.2858360116... | dev |
| 320 | 6844VNI   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/16/2024 1:10:1... | 28.5298954108741   | -16.2858360116... | dev |
| 321 | 8150VZG   | device3            | Cruce Batanes TF... | Las Mercedes     | approach           | 6/16/2024 5:45:5... | 28.5298954108741   | -16.2858360116... | dev |

Figura 3.10: Ejemplo del explorador de Lakehouse.

- **Cuadernos:** Los ingenieros de datos pueden usar el cuaderno para escribir código para leer, transformar y escribir directamente en *Lakehouse* como tablas o carpetas.
- **Canalizaciones:** Los ingenieros de datos pueden usar herramientas de integración de datos como la herramienta de copia de canalización para extraer datos de otros orígenes y copiarlos en *Lakehouse*.

- **Definiciones de trabajos de Apache Spark:** Los ingenieros de datos pueden desarrollar aplicaciones sólidas y organizar la ejecución de trabajos de Spark compilados en Java, Scala y Python.
- **Dataflows Gen 2:** Los ingenieros de datos pueden utilizar Dataflows Gen 2 para ingerir y preparar sus datos.

### 3.7.2 Notebook

Los cuadernos (o bloc de notas) son un entorno de procesamiento interactivo que permite a los usuarios crear y compartir documentos que contienen código activo, ecuaciones, visualizaciones y texto narrativo. Permiten a los usuarios escribir y ejecutar código en varios lenguajes de programación, como Python, R y Scala. Podemos usar cuadernos para la ingesta de datos, preparación, análisis y otras tareas relacionadas con datos.

### 3.7.3 Entorno

La funcionalidad de entorno permite a los ingenieros de datos configurar bibliotecas compartidas, que son colecciones de funciones, y utilidades que pueden ser reutilizadas a lo largo de diferentes proyectos y equipos.

Además, la configuración de procesos de Spark es esencial para optimizar el rendimiento de los trabajos de procesamiento de datos. Esto incluye la gestión de parámetros como el número de ejecutores, la memoria asignada a cada ejecutor, y otros parámetros que afectan el rendimiento de Spark. Ajustar estos parámetros adecuadamente puede resultar en una ejecución más rápida y eficiente de los trabajos de procesamiento de datos.

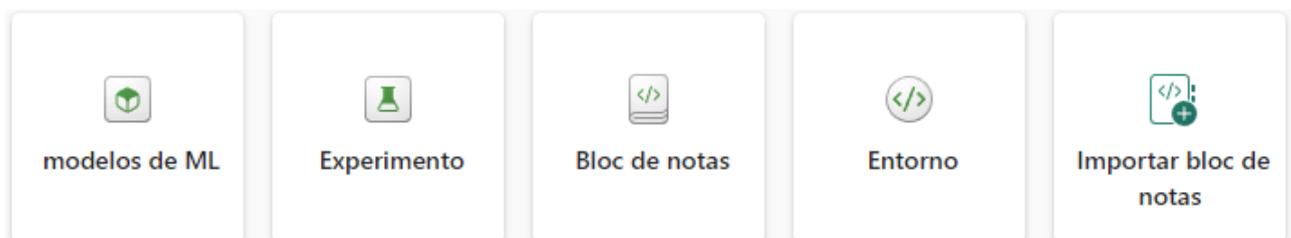
### 3.7.4 Definición de trabajo de Apache Spark

Las definiciones de trabajos de Spark son instrucciones que definen cómo ejecutar un trabajo en un clúster de Spark. Incluyen información como los orígenes de datos de entrada y salida, las transformaciones y los valores de configuración de la aplicación Spark. La definición de trabajo de Spark permite enviar un trabajo por lotes o streaming al clúster de Spark, aplicar una lógica de transformación diferente a los datos hospedados en el *Lakehouse*, además de muchos otros aspectos.

## 3.8 Data Science

Microsoft Fabric ofrece experiencias de ciencia de datos para permitir a los usuarios completar flujos de trabajo de un extremo a otro para el enriquecimiento de datos. Podemos completar una amplia gama de actividades en todo el proceso de ciencia de datos: exploración, preparación y limpieza de los datos, experimentación, modelado, puntuación del modelo y transmisión de información predictiva a los informes de BI.

Los usuarios de Microsoft Fabric pueden acceder a una página principal de ciencia de datos. Desde allí, pueden detectar varios recursos relevantes y acceder a ellos. Por ejemplo, pueden crear experimentos de aprendizaje automático, modelos y cuadernos. También pueden importar cuadernos existentes en la página principal de ciencia de datos.



**Figura 3.11:** *Funcionalidades de Data Science.*

Los usuarios de ciencia de datos en Microsoft Fabric trabajan en la misma plataforma que los usuarios empresariales y los analistas. Como resultado, el uso compartido y la

colaboración de datos se vuelven más transparentes en todos los roles. Los analistas pueden compartir fácilmente informes y conjuntos de datos de Power BI con profesionales de ciencia de datos. La facilidad de colaboración entre roles de Microsoft Fabric hace que las entregas durante la fase de formulación del problema sean mucho más fáciles.

Los usuarios de Microsoft Fabric pueden interactuar con los datos de OneLake mediante el elemento *Lakehouse*. *Lakehouse* se asocia fácilmente a un cuaderno para examinar e interactuar con los datos. Se pueden leer fácilmente datos de una instancia de *Lakehouse* directamente en un dataframe de Pandas. En términos de exploración, esto permite leer datos sin problemas en OneLake.

Existe un potente conjunto de herramientas para la ingesta de datos y las canalizaciones de orquestación de datos con canalizaciones de integración de datos, una parte integrada de forma nativa en Microsoft Fabric. Las canalizaciones de datos (que son fáciles de compilar) pueden acceder a los datos y transformarlos en un formato que el aprendizaje automático pueda usar.

La experiencia de cuadernos de Microsoft Fabric agregó una característica para usar Data Wrangler, una herramienta de código que prepara los datos y genera código de Python. Esta experiencia facilita la aceleración de tareas tediosas como la limpieza de datos y fomenta la repetición y la automatización mediante el código generado.

Con herramientas como PySpark/Python, SparklyR/R, los cuadernos pueden controlar el entrenamiento del modelo de aprendizaje automático.

Los algoritmos y bibliotecas de ML pueden ayudar a entrenar modelos de aprendizaje automático. Las herramientas de administración de bibliotecas pueden instalar estas bibliotecas y algoritmos. Por lo tanto, los usuarios tienen la opción de aprovechar una gran variedad de bibliotecas de aprendizaje automático populares para completar el entrenamiento del modelo de ML en Microsoft Fabric.

Además, las bibliotecas populares como Scikit Learn también permiten desarrollar modelos.

Los experimentos y ejecuciones de MLflow pueden realizar un seguimiento del entrenamiento del modelo de ML. Microsoft Fabric ofrece una experiencia MLflow integrada con la que los usuarios pueden interactuar, para registrar experimentos y modelos.

En Microsoft Fabric, los valores de predicción se pueden escribir fácilmente en OneLake y consumirse sin problemas desde informes de Power BI, con el modo de Direct Lake de Power BI. Esto facilita a los profesionales de la ciencia de datos compartir resultados de su trabajo con las partes interesadas y también simplifica la operacionalización.

## 3.9 Data Warehouse

Microsoft Fabric proporciona a los clientes un producto unificado que aborda todos los aspectos de su patrimonio de datos al ofrecer una plataforma completa de datos SaaS, análisis e inteligencia artificial, centrada en lago y abierta. La base de Microsoft Fabric permite al usuario aprovechar las cargas de trabajo de base de datos, análisis, mensajería, integración de datos e inteligencia empresarial a través de una experiencia SaaS compartida con Microsoft OneLake como pieza central.

Microsoft Fabric presenta un almacenamiento de datos centrado en lago basado en un motor de procesamiento que permite el rendimiento a gran escala. A través de una experiencia SaaS que está estrechamente integrada con Power BI para facilitar el análisis y la creación de informes, Warehouse en Microsoft Fabric converge el mundo de los lagos de datos y los almacenes con el objetivo de simplificar considerablemente la inversión de una organización en su patrimonio de análisis. Las cargas de trabajo de almacenamiento

de datos se benefician de las funcionalidades enriquecidas del motor de SQL sobre un formato de datos abierto, lo que permite a los usuarios centrarse en la preparación, el análisis y los informes de datos en una sola copia de sus datos almacenados en Microsoft OneLake.



Figura 3.12: Funcionalidades de Data Warehouse.

## 3.10 OneLake

Un lago de datos es la base en la que se crean todas las cargas de trabajo de Fabric. Microsoft Fabric Lake también se conoce como OneLake. OneLake está integrado en la plataforma Fabric y proporciona una ubicación unificada para almacenar todos los datos de la organización en los que funcionan las cargas de trabajo.

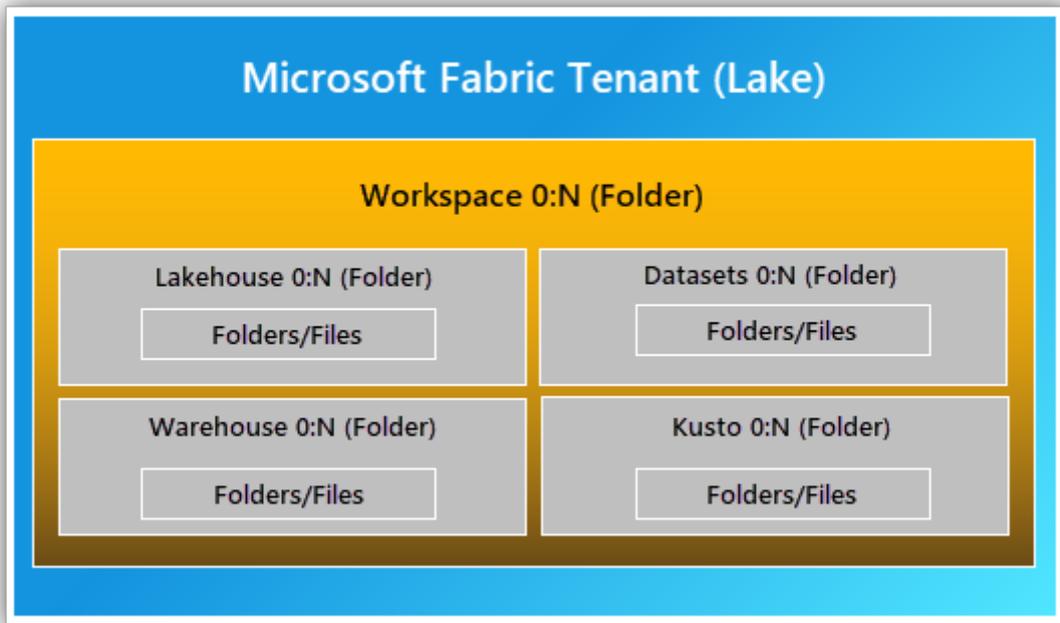
OneLake se basa en ADLS (*Azure Data Lake Storage*) Gen2. Proporciona una única experiencia de SaaS y simplifica las experiencias de Fabric, lo que elimina la necesidad de que se comprenda conceptos de infraestructura, como los grupos de recursos, el RBAC (Access Control basado en roles), Azure Resource Manager, la redundancia o las regiones.

OneLake es jerárquico por naturaleza para simplificar la administración en toda la organización. Microsoft Fabric incluye OneLake y no hay ningún requisito de aprovisionamiento por adelantado. Solo hay una instancia de OneLake por inquilino y proporciona un único espacio de nombre de sistema de archivos que abarca usuarios, regiones y nubes. OneLake organiza los datos en contenedores administrables para facilitar el control.

Un inquilino (*tenant*) es una entidad que representa una organización o empresa que utiliza los servicios de Microsoft. Cada inquilino es un contenedor dedicado que incluye todos los usuarios, grupos, aplicaciones y datos de la organización.

El inquilino se asigna a la raíz de OneLake y está en el nivel superior de la jerarquía. Puede crear cualquier número de áreas de trabajo dentro de un inquilino, las cuales puede considerar como carpetas.

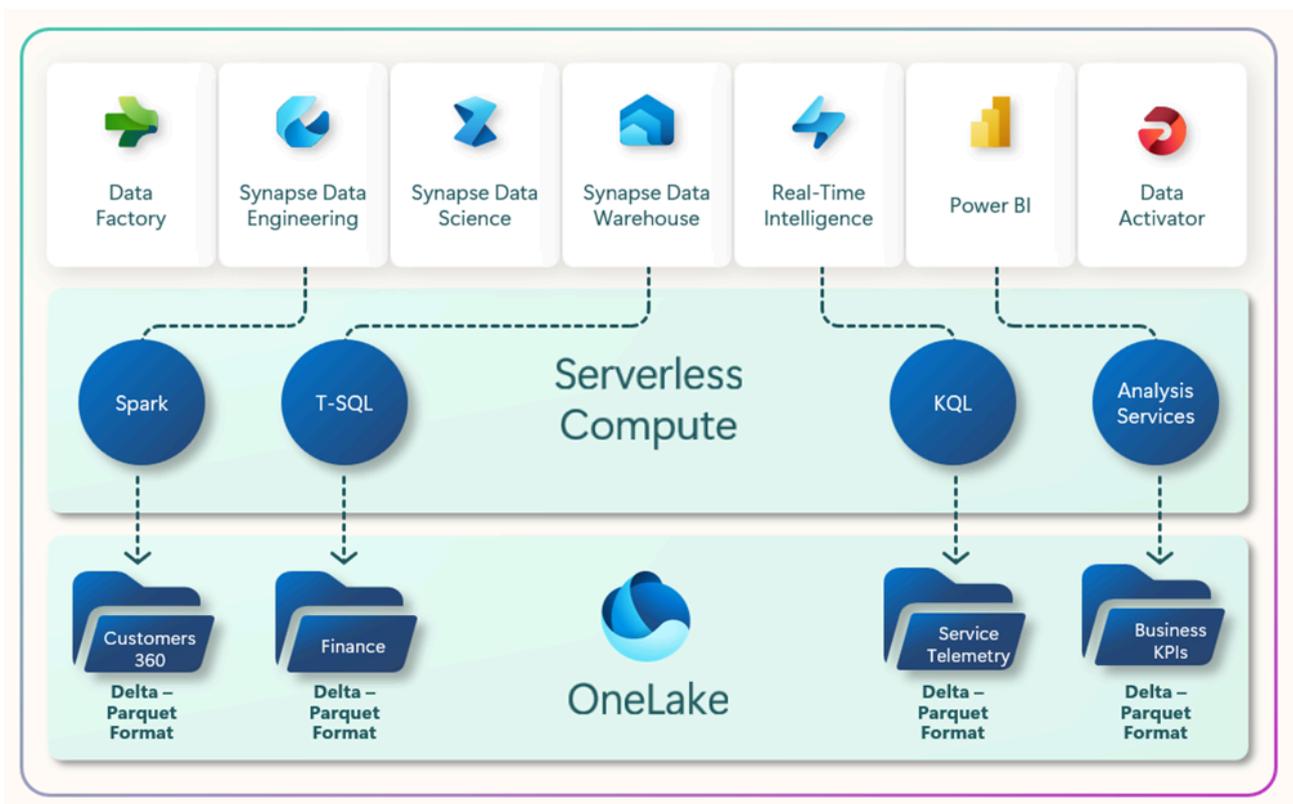
En la siguiente figura se muestra cómo Fabric almacena datos en varios elementos dentro de OneLake. Como se muestra, se pueden crear varias áreas de trabajo dentro de un inquilino y crear varias instancias de almacenes de lago dentro de cada área de trabajo.



**Figura 3.13:** Jerarquía de un inquilino en Microsoft Fabric.

Cada desarrollador puede crear instantáneamente sus propias cargas de trabajo en OneLake. Además, pueden ingerir datos en sus propios almacenes de lago y luego empezar a procesar, analizar y colaborar en los datos.

Todas las experiencias de proceso de Microsoft Fabric están preconectadas a OneLake. Los componentes como ingeniería de datos, Data Warehouse, Data factory, power BI e inteligencia en tiempo real usan OneLake como su almacén nativo. No necesitan ninguna configuración adicional.



**Figura 3.14:** Arquitectura de OneLake.

## 3.11 Flujos de tareas en Microsoft Fabric

Una de las nuevas características añadidas en Microsoft Fabric es la de implementar flujos de tareas. El flujo de tareas de Fabric es una característica de área de trabajo que permite crear una visualización del flujo de trabajo en el área de trabajo. El flujo de tareas ayuda a comprender cómo los elementos están relacionados, funcionan juntos en el área de trabajo y facilitan la navegación por el área de trabajo, incluso cuando esta se vuelve más compleja a lo largo del tiempo.

Fabric proporciona una gama de flujos de tareas predefinidos de un extremo a otro basados en los procedimientos recomendados que están diseñados para facilitar la introducción al proyecto.

Cada área de trabajo tiene un flujo de tareas. El flujo de tareas ocupa la parte superior de la vista de lista del área de trabajo. Consta de un lienzo donde puede crear la visualización del proyecto de análisis de datos y un panel lateral donde se puede ver y editar detalles sobre el flujo de tareas, las tareas y los conectores.

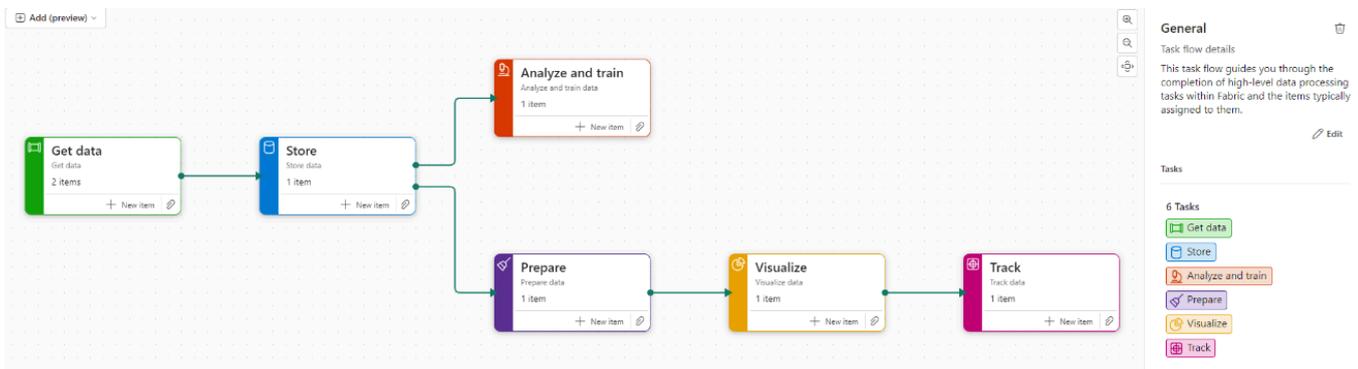


Figura 3.15: Ejemplo de diagrama de tareas.

# Capítulo 4 Prototipo implementado

En este capítulo explicaremos la implementación de un prototipo para el análisis del tráfico y la movilidad en el macizo de Anaga. Este prototipo busca simular la recopilación y análisis de datos de tráfico utilizando información proporcionada por dispositivos como cámaras, que estarán ubicadas en las distintas carreteras de la zona de Anaga, tanto en sus entradas como en sus salidas. Dado que las cámaras aún no han sido instaladas y no contamos con datos reales, la compañera, Alicia Guadalupe Cruz Perez, nos ha proporcionado el código de su Trabajo de Fin de Grado, el cual es capaz de generar un simulador con los datos que se utilizarán en la plataforma de datos.

Gracias al código proporcionado por nuestra compañera, podemos simular el tráfico de vehículos a través de diversas cámaras y rastrear las rutas que siguen. Con esta simulación, es posible generar una matriz origen-destino para cada vehículo, lo que nos permitirá realizar un análisis detallado de estos datos.

Para llevar a cabo este prototipo utilizaremos la herramienta de Microsoft Fabric disponible a través de la licencia de la ULL (<https://app.fabric.microsoft.com/home> con la cuenta institucional, @ull.edu.es). Una vez registrado, pulsaremos sobre nuestra foto de perfil y activaremos la licencia de prueba, la cual nos proporciona 60 días de uso en toda la plataforma. Una vez pasado este periodo, si queremos seguir utilizando la plataforma, debemos crear una capacidad de Fabric. Este proceso se realiza desde el portal de Azure (<https://portal.azure.com/>). Una vez dentro, podemos comprar la capacidad de Fabric y seleccionar el tamaño que se ajuste a nuestro proyecto. Dado que el precio es mensual, podemos empezar con un tamaño menor e ir escalando a tamaños superiores en caso de que la cantidad de datos aumente.

A continuación, detallamos los pasos a seguir para el desarrollo del prototipo en la plataforma de datos Microsoft Fabric.

## 4.1 Crear un área de trabajo

El primer paso en nuestro proyecto es crear una nueva área de trabajo en Microsoft Fabric. Esta área de trabajo servirá como el entorno centralizado donde gestionaremos todos los recursos y procesos necesarios para el análisis de los datos.

Para ello, accedemos al portal de Microsoft Fabric. Navegamos hasta la sección de Áreas de trabajo y seleccionamos "Nueva área de trabajo". Agregamos un nombre descriptivo y relevante para el proyecto.

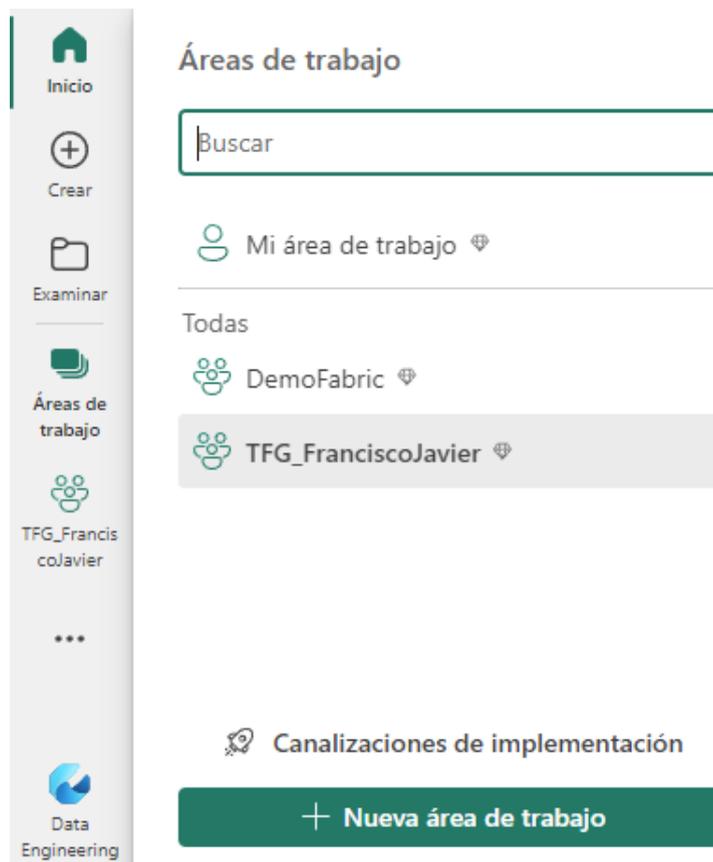


Figura 4.1: Interfaz de las áreas de trabajo en Microsoft Fabric.

## 4.2 Crear un Lakehouse

Una vez creada el área de trabajo, procederemos a crear el *Lakehouse*, el cual será el repositorio principal donde almacenaremos todos los datos generados por el simulador. Esta solución de almacenamiento escalable es ideal para manejar grandes volúmenes de datos.

Dentro del área de trabajo recién creada, seleccionamos la opción para crear un nuevo *Lakehouse*. Definimos la estructura del *Lakehouse*, incluyendo carpetas y subcarpetas para organizar los datos de manera lógica y poder añadir datos tanto estructurados como no estructurados.

## 4.3 Crear una canalización de datos

Tras esto debemos encontrar una forma para introducir los datos directamente al *Lakehouse*. Una forma de hacerlo es simplemente arrastrando el fichero de datos (ya sea un fichero csv, parquet...) al *Lakehouse*, con esto podemos agregar datos de forma sencilla pero a la hora de tener un flujo automatizado no es la mejor opción ya que cada vez que queramos actualizar los datos tenemos que subir el fichero manualmente.

Para arreglar esto lo que vamos a hacer es crear una canalización de datos, con esto vamos a poder cargar los datos desde diversas fuentes y hacer conexiones entre donde se encuentran los datos y Microsoft Fabric.

Una vez creada la canalización y haberle puesto un nombre, seleccionamos la opción de copiar datos y ahí tendremos dos opciones, una para agregarla al lienzo, la cual tendremos que configurar manualmente eligiendo el origen y el destino de los datos a copiar. La segunda opción es el asistente de copia de datos, con el cual nos guiará paso a paso para seleccionar el origen de datos (Carpeta, Base de datos MySQL, Http, REST,

etc.), hacer la conexión con el origen de los datos, seleccionar el destino de los datos (*Lakehouse*, *Warehouse*...) y poder revisar y previsualizar los datos antes de guardar las opciones.

En nuestro caso, como los datos son generados mediante un código ejecutado en local debemos hacer una conexión local mediante una puerta de enlace de datos local (*on-premises data gateway*). Una vez descargada y configurada la puerta de enlace, debemos crear la conexión con fabric (Configuración → Administrar conexiones y puertas de enlace → Nuevo → Local → Nombre del cluster que pusimos en nuestra puerta de enlace en local → Nombre de la Conexión → Crear). Con esto ya podemos seleccionar nuestra conexión desde local y obtener los datos seleccionando la carpeta y el fichero donde se encuentran.

General **Origen** Destino Asignación Configuración

Conexión \*  Actualizar Prueba de conexión Editar

Tipo de ruta de acceso de archivo  Ruta de acceso del archivo  Filtro de archivo  Ruta de acceso de archivo de comodín  Lista de archivos

Ruta de acceso del archivo  /  Examinar | Vista previa de los datos

Recursively

Formato de archivo \*  Configuración

> Avanzadas

**Figura 4.2:** Configuración de origen de datos.

Para el destino de los datos seleccionaremos donde queremos que esos datos se almacenen, en nuestro caso en el *Lakehouse* anteriormente creado, y tenemos la opción de que esos datos se guarden como una tabla ya estructurada o en archivos (*csv*, *parquet*...), ponerle un nombre a la tabla/fichero destino y visualización previa de los datos a cargar. Además, tendremos una opción avanzada para el caso en que guardemos los datos en forma de tabla que nos permitirá poder anexar o sobrescribir los datos importados, así cada vez que ejecutemos la canalización de datos podremos añadir los datos extraídos a la tabla o borrarlos y añadir solo los nuevos.

General Origen **Destino** Asignación Configuración

Conexión \*  Actualizar Abierto

Carpeta raíz  Tablas  Archivos

Nombre de tabla  Actualizar Vista previa de los datos + Nuevo

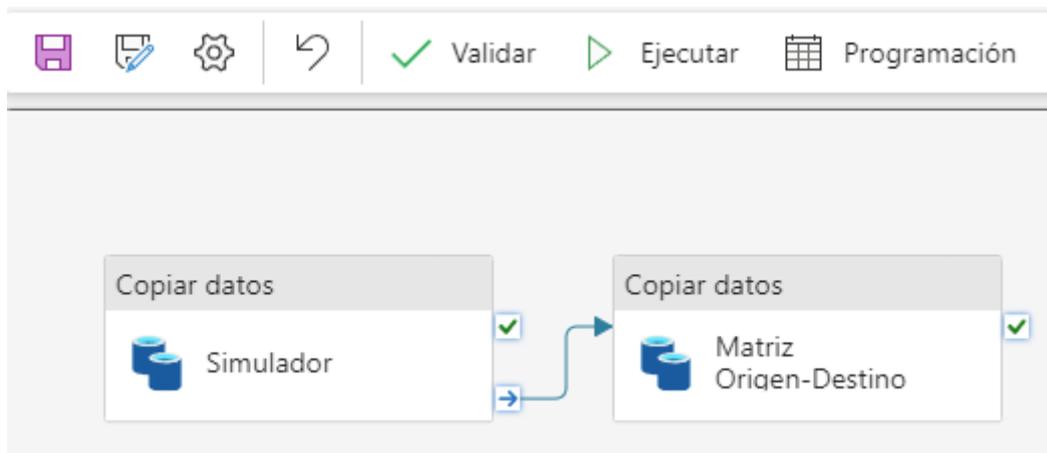
Avanzadas

Acción de tabla  Anexar  Sobrescribir

Columnas de partición No se ha detectado ninguna columna de partición en la tabla actual de Lakehouse.

**Figura 4.3:** Configuración de destino de datos.

Con esto hecho ya podremos extraer los datos del simulador y la matriz origen-destino de la carpeta local y, al añadirlos a la canalización de datos, quedará ya configurado la forma de extracción de los datos y podremos ejecutarlo siempre que queramos dándole al botón de ejecutar.

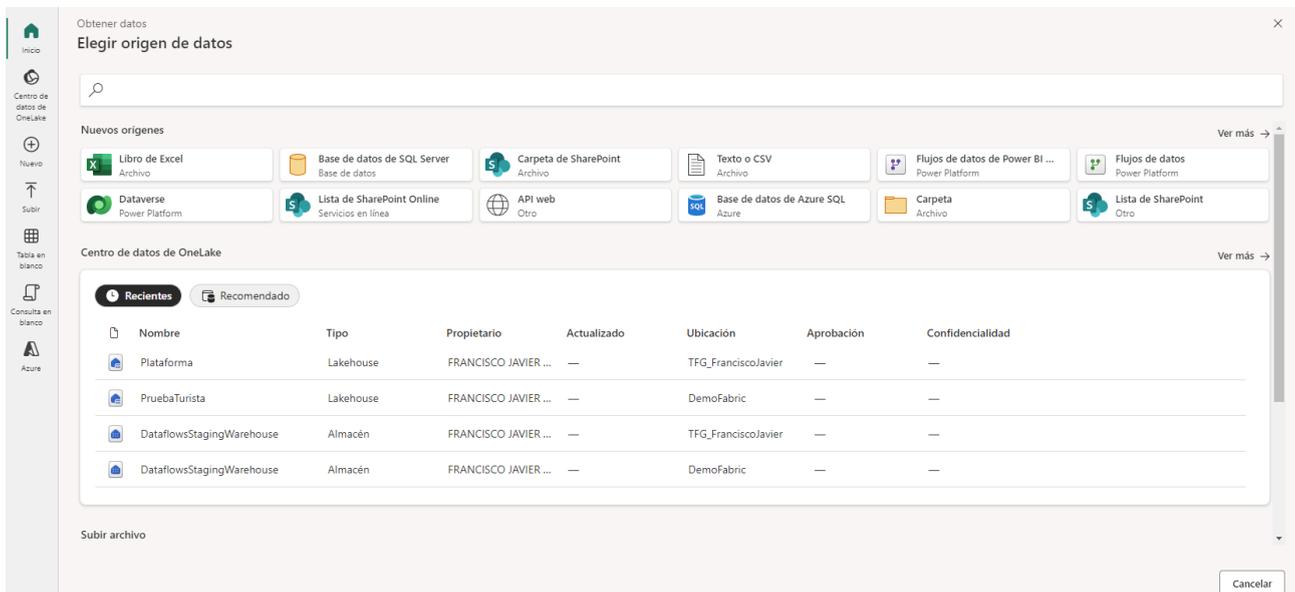


**Figura 4.4:** Canalizador de datos con la copia de datos.

## 4.4 Crear un flujo de datos

Tras esto, el siguiente paso será crear un flujo de datos Gen2 para modificar la estructura de los datos de matriz origen-destino, esto se hace con la intención de ajustar aspectos como los nombres de las columnas, el tipo de las variables (fecha, cadena, entero, decimal...), reemplazar caracteres y agregar/quitar/combinar columnas.

En nuestro caso, debemos darle al botón de “*Obtener datos*” y en “*Más...*” aparecerá nuestro centro de datos de *OneLake* y podremos elegir los datos de la matriz origen-destino almacenados en nuestro Lakehouse para cargar la tabla al flujo de datos.



**Figura 4.5:** Obtener datos desde un flujo de datos Gen2.

Una vez cargados los datos al flujo aparecerá una interfaz con la cual podremos interactuar con las columnas, ya sea cambiando el nombre pulsando dos veces encima del nombre de la columna que queremos cambiar, como el tipo de variable pulsando encima del valor que aparece al lado del nombre de la columna.

|    | Matrícula                                 | Dispositivo_Entrada | Carretera_Entrada          | Zona_Entrada   | Direccion_entrada | Fecha_Hora_Entrada |
|----|---|---------------------|----------------------------|----------------|-------------------|--------------------|
| 1  | 1.2 Número decimal                        |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 0:27:43  |
| 2  | \$ Divisa                                 |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 0:35:50  |
| 3  | 1 <sup>2</sup> 3 Número entero            |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 0:54:20  |
| 4  | % Porcentaje                              |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 2:27:44  |
| 5  | 📅 Fecha/Hora                              |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 1:04:30  |
| 6  | 📅 Fecha                                   |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 4:14:18  |
| 7  | 🕒 Tiempo                                  |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 5:30:37  |
| 8  | 🕒 Tiempo                                  |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 6:14:07  |
| 9  | 🌐 Fecha/hora/zona                         |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 6:07:04  |
| 10 | 🕒 Duración                                |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 6:25:26  |
| 11 | 📄 Texto                                   |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 8:16:34  |
| 12 | ✓ Verdadero/Falso                         |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 8:21:24  |
| 13 |   |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 10:11:06 |
| 14 | 010<br>101 Binario                        |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 9:33:26  |
| 15 | ABC<br>123 Usar configuración regional... |                     | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 8:57:02  |
| 16 | 124URJZ                                   | device2             | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 9:21:55  |
| 17 | 2903QGW                                   | device2             | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 10:24:55 |
| 18 | 7530LTB                                   | device2             | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 10:19:59 |
| 19 | 3071GRK                                   | device2             | Cruce Bejia/Batanes TF-143 | Bejias-Batanes | approach          | 1/7/2024, 11:40:46 |

Figura 4.6: Cambio de tipo de la columna.

Dentro de las distintas transformaciones que tuvimos que hacer (cambio de nombre, cambio de tipo y cambio de caracteres), una de las transformaciones necesarias fue la de combinar la columna de Fecha y Hora, esto fue debido a que mientras que la columna Fecha si era posible ser de tipo Fecha, la columna Hora no era posible ser de tipo Tiempo puesto que la herramienta actualmente no admite el tipo. Para solucionar esto debemos ir a la pestaña de "Agregar columna" y en columna personalizada añadiremos un pequeño script hecho mediante el lenguaje de fórmulas M de Power Query. Con esto podremos hacer la fecha y la hora tanto de entrada como de salida de vehículos y podremos eliminar las otras columnas para no tener redundancia de datos.

### Columna personalizada ?

Agregue una columna que se calcula a partir de otras columnas o valores.

Nuevo nombre de columna \*

Tipo de datos

Fecha/Hora

Fórmula de columna personalizada \* ?

```
= #datetime(Date.Year([Fecha_Entrada]), Date.Month([Fecha_Entrada]), Date.Day([Fecha_Entrada]), Time.Hour([Hora_Entrada]), Time.Minute([Hora_Entrada]), Time.Second([Hora_Entrada]))
```

Columnas disponibles

- Matrícula
- Dispositivo\_Entrada
- Carretera\_Entrada
- Zona\_Entrada
- Direccion\_entrada
- Fecha\_Entrada
- Hora\_Entrada

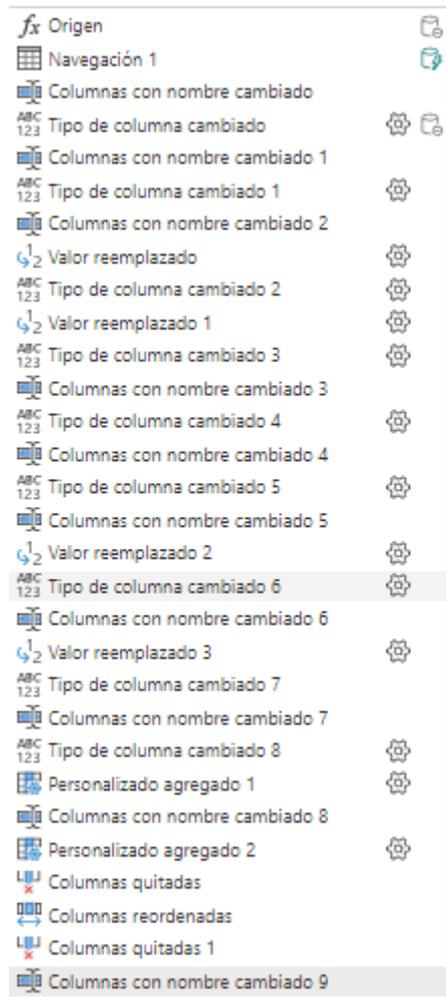
Insertar columna

[Más información sobre las fórmulas de Power Query](#)

Aceptar Cancelar

Figura 4.7: Script en M que combina las columnas Fecha y Hora.

Hay que tener en cuenta que cada paso que realicemos en el flujo de datos se guardará como un paso aplicado dentro del flujo, así quedarán todos los pasos registrados al terminar un flujo y cuando lo ejecutemos, se van a procesar las columnas de la tabla de la misma forma en que lo hicimos dentro del flujo.



**Figura 4.8:** Pasos aplicados dentro del flujo de datos.

Una vez hemos transformado la tabla con los datos a nuestra necesidad, debemos elegir el destino de estos nuevos datos, para ello seleccionamos en la pestaña de “Inicio” la opción de “Agregar destino de datos” y seleccionamos el Lakehouse creado, ahí podremos darle un nombre a la nueva tabla y después de hacer esto le damos al botón “Publicar”.

Con el flujo de datos publicado y creado dentro de nuestra área de trabajo, vamos a añadirlo a la canalización de datos. Dentro de la canalización pulsaremos sobre *flujo de datos* y haremos la configuración añadiendo el área de trabajo y el flujo de datos. Con esto lo conectaremos desde la copia de los datos para obtener un flujo de trabajo automatizado.

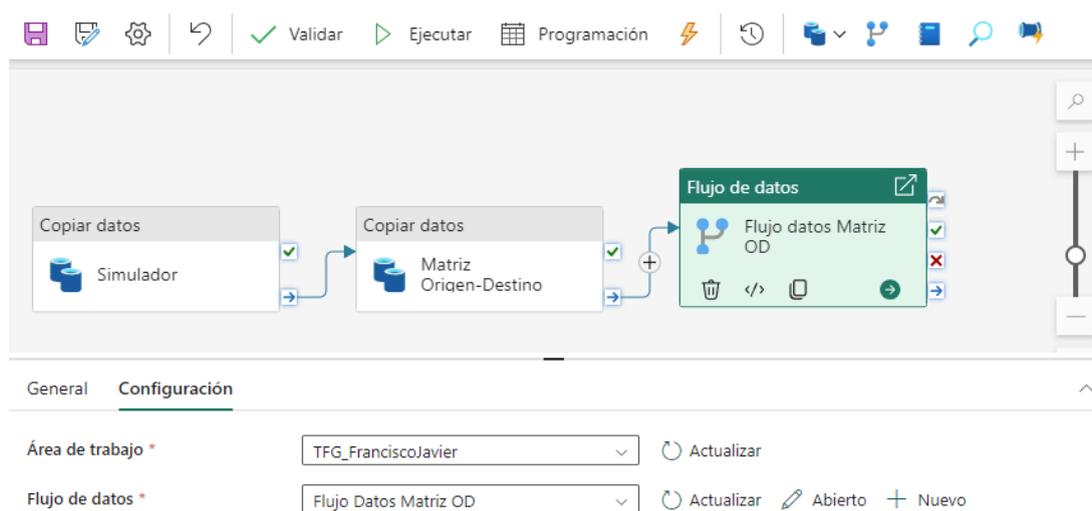


Figura 4.9: Flujo de datos configurado dentro de la canalización.

## 4.5 Crear un modelo semántico

Ahora, debemos crear un modelo semántico que va a actuar como una capa de abstracción sobre los datos almacenados en nuestro Lakehouse. Con este modelo no solo podemos organizar los datos de manera estructurada, sino que también podemos establecer las relaciones entre diferentes conjuntos de datos, lo que permite hacer un análisis más profundo y significativo.

Dentro de cada *Lakehouse* hay asociado un modelo semántico con el que trabajar, sin embargo, es posible crear tantos modelos semánticos como queramos y añadir los distintos conjuntos de datos en base a los datos con los que se van a trabajar. Para nuestro caso crearemos un modelo semántico que después sea capaz de actualizar los datos desde la canalización, esto es debido a que el modelo semántico asociado a nuestro *Lakehouse* no puede actualizarse automáticamente.

Para crear el modelo semántico lo único que tendremos que hacer es entrar en nuestro *Lakehouse* y una vez dentro seleccionamos en “Nuevo modelo semántico” y ahí seleccionamos los conjuntos con los que queremos trabajar.

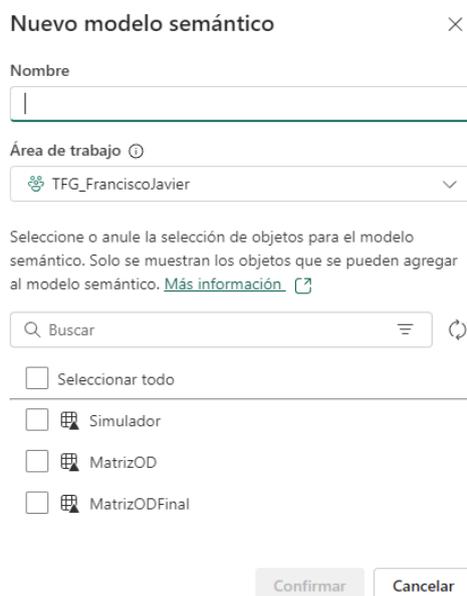


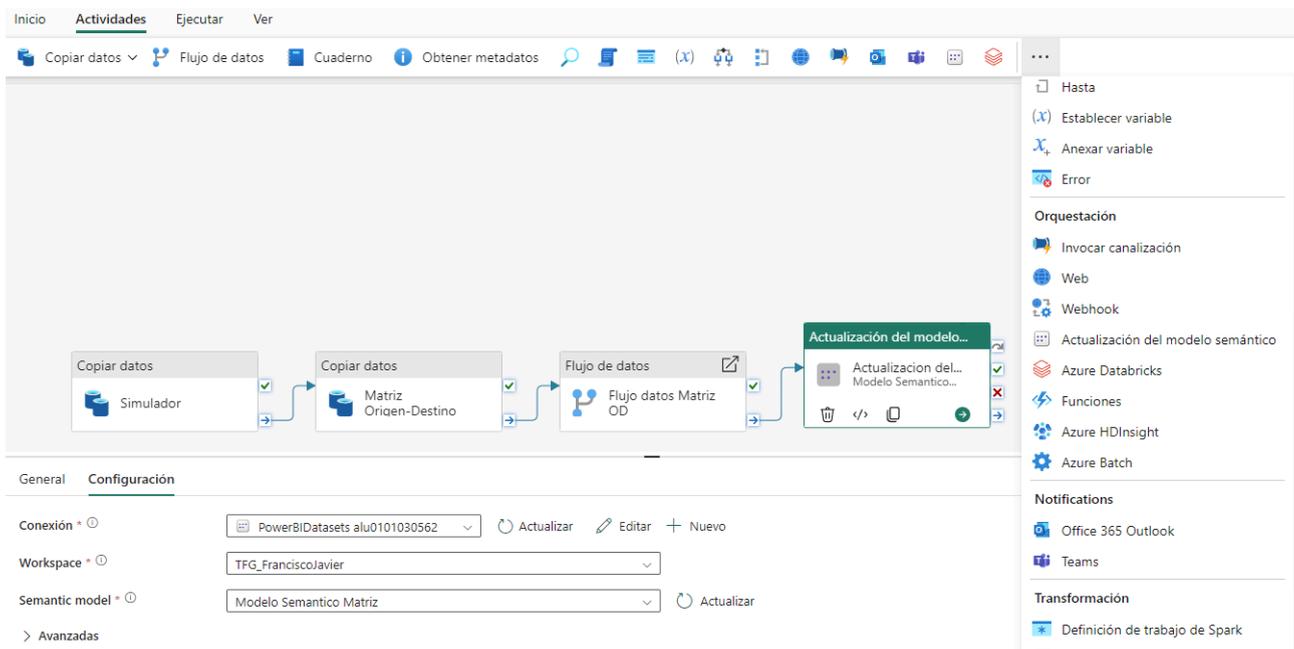
Figura 4.10: Creación de un modelo semántico.

Una vez definido el modelo, configuramos las relaciones entre las tablas identificando campos clave que las conectan, como identificadores únicos o fechas de eventos. Estas relaciones permiten a Fabric entender cómo están relacionados los datos entre diferentes conjuntos. En nuestro caso, la matriz origen-destino transformada en el flujo de datos será suficiente, ya que contiene toda la información necesaria para realizar un análisis en profundidad con la herramienta Power BI.

Es importante tener en cuenta que cuando se trabaja con grandes volúmenes de datos distribuidos en múltiples tablas, tenemos que considerar cómo estas tablas interactúan entre sí y cómo impactan en el rendimiento del sistema. La computación dentro de una herramienta como Power BI implica procesar y manipular estos datos para presentarlos de manera efectiva.

Por ello cuantas más tablas haya y más complejas sean las relaciones entre ellas, mayor será la carga computacional al ejecutar consultas. Power BI necesita procesar estas relaciones y extraer datos de múltiples fuentes, lo que puede afectar la velocidad de respuesta y el tiempo de carga de los informes. Por eso, un modelo de datos bien diseñado puede optimizar el rendimiento al minimizar la cantidad de datos redundantes y asegurar que las relaciones entre tablas estén bien definidas.

Tras haber creado el modelo semántico, lo único que falta es añadirlo a la canalización para que, una vez lo ejecutemos, se haga todo el proceso automáticamente, desde la carga de datos, la transformación y la actualización del modelo semántico. Para ello dentro de la canalización iremos a la pestaña *Actividades* y pulsaremos sobre el botón de “Actualización del modelo semántico”. Una vez añadido lo configuramos con nuestra área de trabajo y el modelo semántico asociado.



**Figura 4.11:** Modelo semántico configurado en la canalización de datos.

## 4.6 Crear un informe en Power BI

Con esto, ya podemos centrarnos en la creación de un informe en Power BI. Para ello tenemos dos opciones, crear un informe desde el modelo semántico en la página de Microsoft Fabric o desde la aplicación de escritorio de Power BI. En este último caso, podemos vincular nuestro OneLake autenticándonos en la aplicación con el mismo correo electrónico que en Microsoft Fabric. Para nuestro caso, usaremos la aplicación de escritorio, ya que nos permite más flexibilidad a la hora de realizar medidas en DAX y

hacer uso de Direct Query.

DAX (Data Analysis Expressions) es un lenguaje de fórmulas que se usa en Power BI para crear cálculos personalizados en los modelos de datos. Las medidas en DAX son cálculos dinámicos que se realizan sobre los datos a medida que los usuarios interactúan con los informes. Estos cálculos pueden incluir sumas, promedios y más operaciones matemáticas o incluso filtros de selección. Las medidas son esenciales para el análisis de datos porque permiten obtener insights específicos y personalizados según las necesidades del análisis. Al definir una medida en DAX, podemos utilizar funciones como SUM, ISFILTERED, CONCATENATEX y muchas otras, combinadas con lógica condicional y operadores matemáticos para crear cálculos que se adapten a cualquier contexto de filtrado aplicado en los informes.

Direct Query es una característica de Power BI que permite a los usuarios conectarse directamente a las bases de datos y consultar datos en tiempo real. A diferencia del modo de importación, donde los datos se cargan y almacenan en Power BI, Direct Query mantiene la conexión con la fuente de datos y ejecuta consultas directamente sobre ella cada vez que se visualizan o interactúan con los datos en el informe. Esto es especialmente útil para trabajar con grandes volúmenes de datos o cuando es necesario disponer de la información más actualizada en todo momento. Sin embargo, el uso de Direct Query puede tener implicaciones en el rendimiento y la velocidad de respuesta de los informes, ya que cada interacción genera consultas a la base de datos. Es crucial optimizar tanto las consultas como la estructura del modelo de datos para asegurar un rendimiento eficiente.

Al abrir la aplicación de escritorio de Power BI debemos iniciar sesión con la misma cuenta con la que estamos trabajando en Microsoft Fabric. Una vez iniciada la sesión, podremos crear un informe en blanco o directamente seleccionar la opción de “Centro de datos de OneLake”. Con esto podremos directamente seleccionar nuestro modelo semántico el cual traerá todas las tablas asociadas.



**Figura 4.12:** Selección de origen de datos desde Power BI desktop.

Tras haber importado el modelo semántico a la aplicación, ya podemos trabajar creando medidas y gráficas que nos servirán para explorar y comunicar los datos de manera efectiva. Las gráficas en Power BI desempeñan un papel crucial al transformar datos complejos en representaciones visuales comprensibles. Permiten identificar patrones y tendencias, proporcionando insights valiosos que facilitan la toma de decisiones informadas.

En nuestro caso, hemos realizado un informe que consta de cuatro pestañas distintas. La primera de estas pestañas actúa como un menú interactivo desde donde se puede acceder fácilmente a las otras tres pestañas, cada una de las cuales ofrece información diferente para explorar.



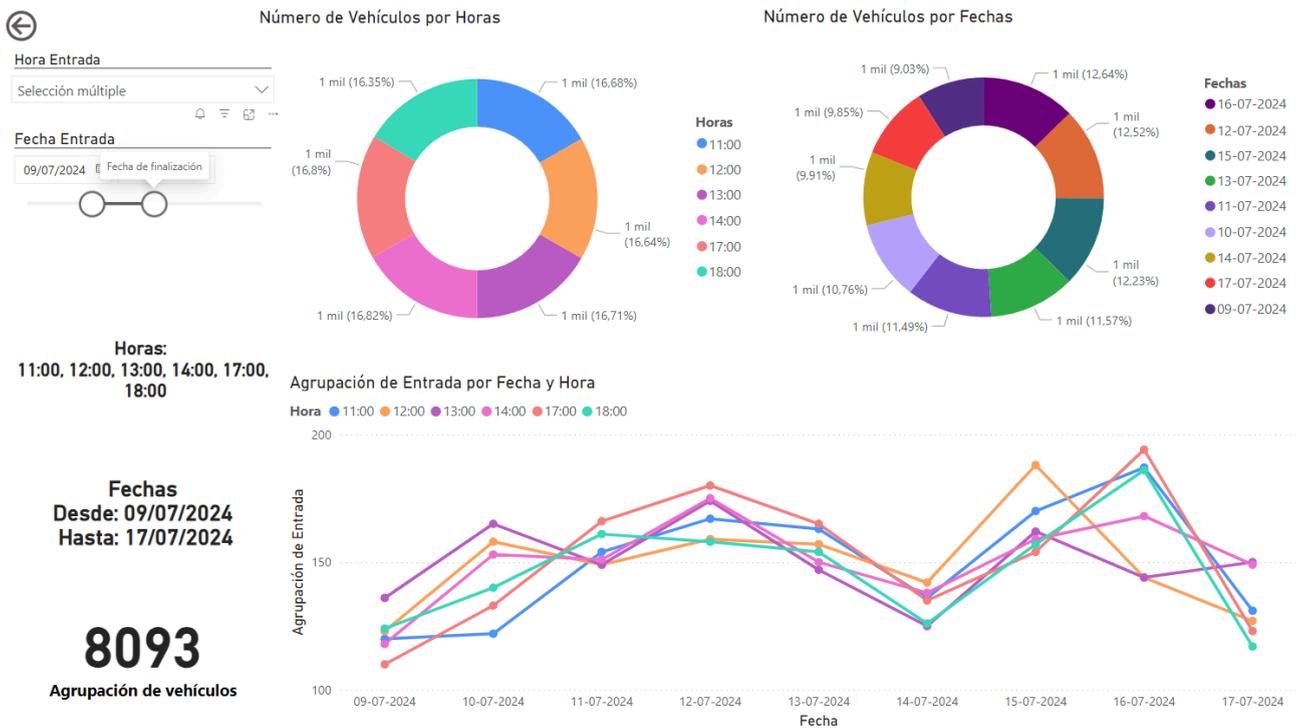
**Figura 4.13:** *Pestaña de menú.*

La primera pestaña de gráficos proporciona un análisis detallado del número de vehículos que entran en la zona de Anaga, segmentado por horas y fechas durante un rango de fechas. La visualización incluye gráficos de anillos, un gráfico de líneas y filtros interactivos.

En cuanto a los filtros interactivos consta de un filtro para la hora de entrada, la cual permite seleccionar distintas horas o visualizar los datos de todas las horas. Para el filtro de fechas de entrada, este permite al usuario seleccionar un periodo de tiempo específico para analizar el tráfico.

Para la parte de los gráficos tenemos dos tipos, el gráfico de anillos, el cual consta del número de vehículos por hora que muestra la distribución de vehículos por cada hora del día seleccionada en el filtro. Cada segmento del anillo representa una hora específica, y el tamaño del segmento indica el porcentaje del total de vehículos que ingresaron durante esa hora, y del número de vehículos por fechas, el cual tiene la misma estructura que el de por horas, solo que visualizando los días seleccionados en el filtro del rango de fechas. El otro tipo de gráfica, el gráfico de líneas, muestra una visión combinada de la cantidad de vehículos que ingresaron por fecha y hora. Cada línea de color representa una hora específica del día, permitiendo visualizar tendencias diarias y patrones de tráfico.

Como etiquetas tenemos las asociadas a los filtros, las cuales nos permiten visualizar las opciones escogidas por el usuario y el número total de vehículos que ingresaron en la zona de Managua durante las horas y las fechas seleccionadas.



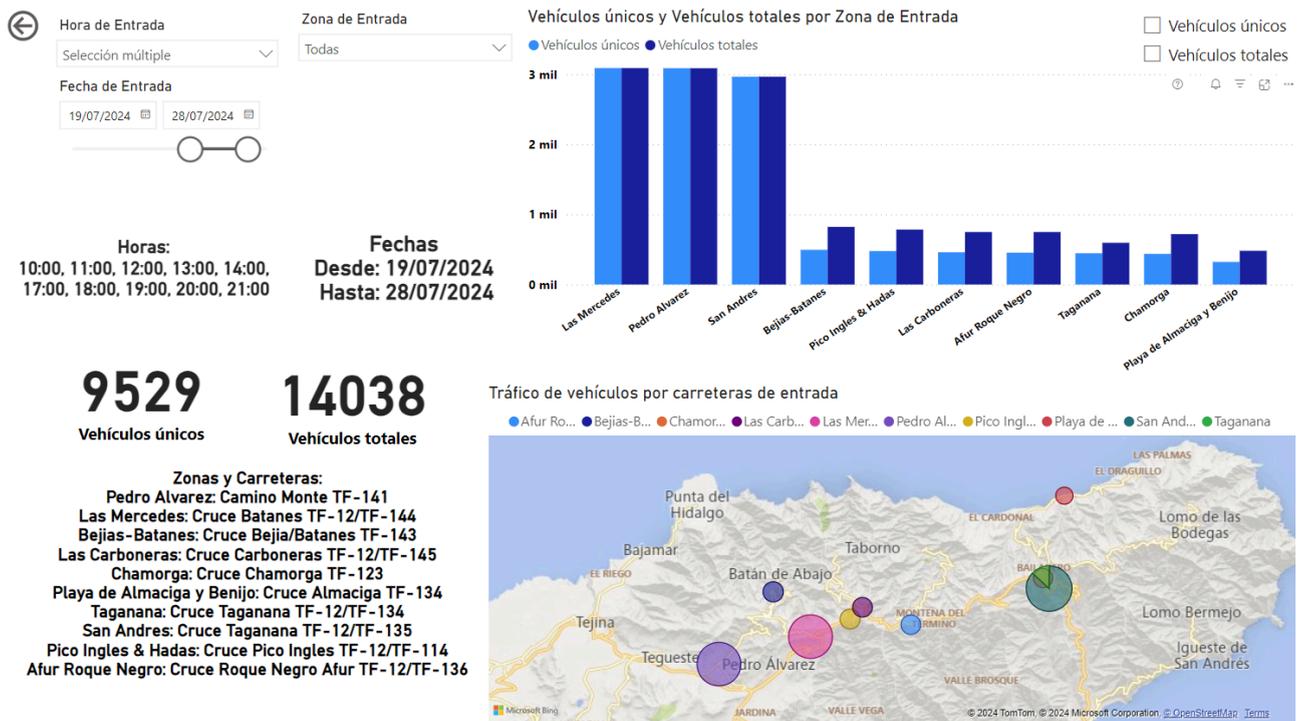
**Figura 4.14:** Pestaña de agrupación de vehículos de entrada por fecha y hora.

La siguiente pestaña está diseñada para mostrar el tráfico de vehículos por zonas de entrada a Anaga. Se utilizan varias visualizaciones para ofrecer una visión completa del número de vehículos únicos y totales que ingresan a través de diferentes carreteras.

En cuanto a los filtros, la hora de entrada y fecha de entrada son iguales a la pestaña anterior, pudiendo elegir horas en concreto y un rango de fechas determinado. Adicionalmente, se ha añadido un filtro con las diferentes zonas de entrada en la zona de Anaga, pudiendo elegir el usuario cualquiera de ellas.

Para esta pestaña tenemos el gráfico de barras, que muestra la cantidad de vehículos tanto únicos como totales de las diferentes zonas de entrada de Anaga. Además tiene un parámetro asociado por el cual podemos elegir la opción de mostrar solo los vehículos únicos o los vehículos totales. Adicionalmente, se ha añadido un mapa interactivo que muestra la ubicación de las distintas zonas en base a la latitud y a la longitud, teniendo un tamaño de burbuja asociado al número de vehículos, por lo que contra más grande es esta, mayor número de vehículos han entrado por esa zona.

Dentro de las etiquetas tenemos la de los filtros de las horas y rango de fechas elegidas por el usuario. Para la etiqueta de zonas y carreteras, esta se muestra en base a las zonas elegidas por el usuario, mostrando no solo la zona de entrada, sino que también muestra la carretera de entrada asociada a dicha zona. Adicionalmente, se han añadido dos etiquetas con el número total de vehículos únicos y el número total de vehículos.



**Figura 4.15:** Pestaña de tráfico de vehículos por zonas de entrada.

La última pestaña tiene el diseño de mostrar el tráfico de vehículos de entrada a la zona de Anaga segmentado por días de la semana. La visualización incluye gráficos circulares, de líneas y un mapa para proporcionar una visión comprensiva del tráfico.

Para los filtros implementados, el de fecha de entrada y zona de entrada es igual al de la pestaña anterior. Los nuevos filtros tienen la opción de elegir un rango de horas en lugar de seleccionarlás una a una y un desplegable con los diferentes días de la semana para filtrar por los días que el usuario elija.

En cuanto a los gráficos, tenemos dos circulares, el primero se encarga de visualizar la cantidad de vehículos que circulan por día de la semana, mientras que el segundo se encarga de mostrar los vehículos por las diferentes zonas de entradas elegidas en el filtro. El gráfico de líneas se encarga de visualizar la cantidad de vehículos de entrada por las fechas seleccionados en el filtro de rango de fechas mientras que el mapa muestra el número de vehículos en las distintas zonas de entrada, que son seleccionadas mediante el filtro.

En las etiquetas se muestran las distintas opciones elegidas en el filtro, estas son los días de la semana, las zonas de entrada y el rango de fechas y horas seleccionadas. Adicionalmente se ha añadido la etiqueta con el total de vehículos que se actualiza siempre que se van cambiando los distintos filtros.



**Figura 4.16:** Pestaña de tráfico de vehículos por días de la semana.

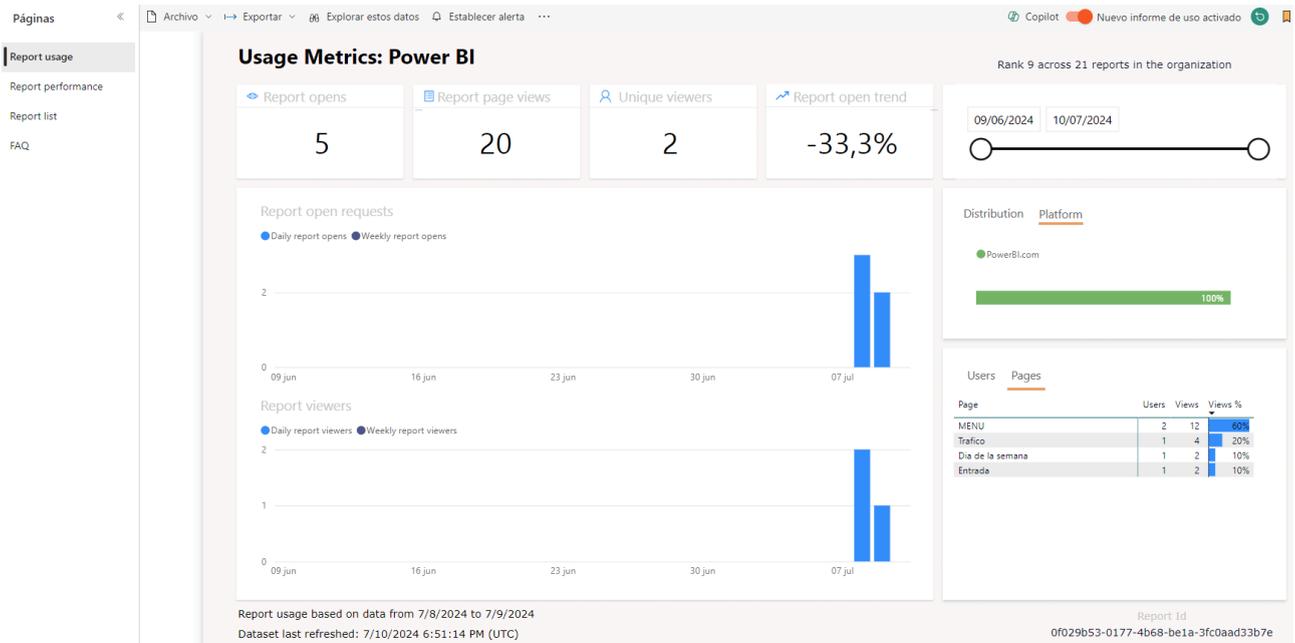
Ahora que hemos finalizado la creación y la configuración del informe en la aplicación de Power BI, podemos obtener una visión clara y detallada del tráfico en la zona de Anaga. A pesar de que los datos utilizados provienen de una simulación y no representan datos reales, las gráficas y las visualizaciones nos permiten identificar patrones y tendencias que serían esenciales en un escenario real.

El siguiente paso será publicarlo en nuestra zona de trabajo en Microsoft Fabric para que esté disponible para su uso y análisis continuo. Para ello debemos desplazarnos hasta la pestaña *inicio*, y seleccionamos “Publicar”, ahí tendremos que escoger nuestra área de trabajo y cargarlo.

Una vez que el informe se haya cargado, podremos acceder a él desde el servicio de Power BI en Microsoft Fabric. Desde allí, tendremos la posibilidad de aplicar diversos filtros y navegar entre las distintas pestañas del informe según sea necesario. Publicar el informe en Microsoft Fabric garantiza que todos los interesados puedan acceder a la información de manera centralizada, permitiendo un análisis más colaborativo.

## 4.7 Monitorización de métricas de uso

Otra opción que ofrece Power BI es la de poder monitorizar las métricas de uso de informes, lo cual es esencial para entender cómo y cuándo se está utilizando el informe por distintos usuarios. Para acceder a estas métricas debemos posicionarnos sobre el informe de power BI y seleccionar los tres puntos, ahí seleccionaremos “Ver informe de métricas de uso” y se nos abrirá el informe con las métricas. Aquí, podremos ver datos detallados sobre la frecuencia de acceso, las páginas más vistas y las interacciones más comunes con el informe.



**Figura 4.17: Informe de métricas de uso.**

## 4.8 Diagrama de tareas

Ahora crearemos el diagrama de tareas que servirá como una guía para cualquier persona que se una al área de trabajo. Este diagrama detalla los pasos y procedimientos necesarios para llevar a cabo las diferentes tareas del proyecto, asegurando que los nuevos integrantes puedan integrarse y entender la estructura del área de trabajo.

En la siguiente Figura se muestra el diagrama de tareas del área de trabajo y como se puede apreciar el flujo empieza obteniendo los datos y una vez obtenidos se guardan en el *Lakehouse* y tras esto, se preparan en el flujo de datos y se guarda la tabla resultante de nuevo en el *Lakehouse*. Tras esto se realiza la parte de visualización con el modelo semántico y el informe en Power BI.



**Figura 4.18: Flujo de tareas del área de trabajo en Microsoft Fabric.**

# Capítulo 5 Conclusiones y líneas futuras

## 5.1 Conclusiones

A lo largo de este proyecto, se han cumplido los objetivos planteados inicialmente, logrando la implementación de un prototipo funcional para el análisis del tráfico en el macizo de Anaga. Utilizando datos simulados proporcionados por código, hemos sido capaces de configurar un sistema completo que abarca desde la recolección y almacenamiento de datos en Microsoft Fabric hasta la visualización de los resultados en informes detallados mediante Power BI. Este proceso ha demostrado que Fabric es un herramienta muy práctica y versátil para el análisis de datos, concretamente para el análisis del tráfico y la movilidad, ofreciendo *insights* valiosos que pueden ser utilizados para mejorar la gestión del tráfico en áreas turísticas.

Además, es importante destacar que Microsoft Fabric está en constante evolución, añadiendo nuevas funcionalidades y métodos para la obtención de datos. Esto significa que herramientas y capacidades que actualmente no se encuentran disponibles podrían implementarse en un futuro cercano, mejorando aún más la funcionalidad y eficiencia de la plataforma.

Sin embargo, otra cosa a destacar es que los costos asociados con el uso de Microsoft Fabric podrían ser elevados, dependiendo de la cantidad de datos y la cantidad de consultas a realizar, lo que podría limitar su adopción en algunos escenarios. Por ello, es importante explorar alternativas open-source que puedan ofrecer capacidades similares, aunque tal vez requieran de un conocimiento mayor y ofrezcan una funcionalidad inferior. Entre las opciones open-source más destacadas se encuentran Apache Hadoop para el almacenamiento y procesamiento de grandes volúmenes de datos [7], Apache Airflow para generar un flujo de trabajo automatizado [8] y Apache Superset para las visualizaciones [9]. Aunque estas alternativas pueden requerir una mayor inversión en términos de configuración y mantenimiento, representan una opción viable para organizaciones con presupuestos más limitados.

## 5.2 Líneas futuras

De cara al futuro, una vez que los dispositivos para la recolección de datos en tiempo real estén plenamente implementados en el macizo de Anaga, sería posible utilizar la misma infraestructura y configuración de informes desarrollados durante este proyecto para analizar los datos reales. Esto permitirá validar las simulaciones realizadas y ajustar los modelos según sea necesario, proporcionando una visión precisa y actualizada del tráfico en la zona. Además, la integración de datos en tiempo real permitirá la implementación de sistemas de alerta y recomendaciones en tiempo real.

Otra adición podría ser la de ampliar el prototipo para incluir el análisis de otros aspectos de la movilidad urbana y rural, como el impacto de eventos específicos o cambios en la infraestructura.

# Capítulo 6 Conclusions and future development

## 6.1 Conclusions

Throughout this project, the initial objectives have been met, achieving the implementation of a functional prototype for traffic analysis in the Anaga massif. Using simulated data provided by code, we have been able to configure a complete system ranging from data collection and storage in Microsoft Fabric to the visualization of the results in detailed reports using Power BI. This process has demonstrated that Fabric is a very practical and versatile tool for data analysis, specifically for traffic and mobility analysis, providing valuable insights that can be used to improve traffic management in tourist areas.

In addition, it is important to note that Microsoft Fabric is constantly evolving, adding new functionalities and methods for obtaining data. This means that tools and capabilities that are currently unavailable could be implemented in the near future, further improving the functionality and efficiency of the platform.

However, another thing to note is that the costs associated with using Microsoft Fabric could be high, depending on the amount of data and the number of queries to be performed, which could limit its adoption in some scenarios. Therefore, it is important to explore open-source alternatives that can offer similar capabilities, although they may require more knowledge and offer inferior functionality. Notable open-source options include Apache Hadoop for storing and processing large volumes of data [\[7\]](#), Apache Airflow for generating an automated workflow [\[8\]](#) and Apache Superset for visualizations [\[9\]](#). Although these alternatives may require a higher investment in terms of configuration and maintenance, they represent a viable option for organizations with more limited budgets.

## 6.2 Future development

Looking ahead, once the real-time data collection devices are fully implemented in the Anaga massif, it would be possible to use the same infrastructure and reporting configuration developed during this project to analyze real data. This will allow validating the simulations performed and adjusting the models as needed, providing an accurate and updated view of the traffic in the area. In addition, the integration of real-time data will allow the implementation of real-time warning systems and recommendations.

Another addition could be to extend the prototype to include analysis of other aspects of urban and rural mobility, such as the impact of specific events or changes in infrastructure.

# Capítulo 7 Presupuesto

A continuación se presenta un presupuesto estimativo del coste de desarrollar este aplicativo. Los costes relativos a la licencia, como hemos mencionado en la memoria, dependerá de los recursos que se necesiten y tendrán que ser estimados con el cliente final.

El tiempo estimado de desarrollo es de **3 meses** y el presupuesto total asciende a **17.100€**. En la siguiente tabla se muestra desglosado el presupuesto.

| Tipos                     | Descripción                  |
|---------------------------|------------------------------|
| Ordenador                 | 600€                         |
| Licencia Microsoft Fabric | 0 € educativa 2 meses        |
| Licencia Microsoft Fabric | Aprox. 500€ / mes<br>x 1 mes |
| Analista                  | 2.000€ al mes                |
| Programador               | 1.500€ al mes                |
| Diseñador                 | 1.500€ al mes                |

**Tabla 7.1:** Coste de desarrollo

# Bibliografía

[1] Qué es Microsoft Azure.

<https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-azure>,  
visitado el 1 de julio de 2024.

[2] Calculadora de precios del servicio Azure.

<https://azure.microsoft.com/es-es/pricing/calculator/>,  
visitado el 1 de julio de 2024.

[3] Qué es Amazon Web Service (AWS).

<https://aws.amazon.com/es/what-is-aws/>,  
visitado el 1 de julio de 2024.

[4] Calculadora de precios del servicio AWS.

[https://calculator.aws/#/?nc2=h\\_ql\\_pr\\_calc](https://calculator.aws/#/?nc2=h_ql_pr_calc),  
visitado el 1 de julio de 2024.

[5] Qué es Google Cloud Platform (GCP),

<https://cloud.google.com/>,  
visitado el 1 de julio de 2024.

[6] Calculadora de precios del servicio GCP.

<https://cloud.google.com/products/calculator/>,  
visitado el 1 de julio de 2024.

[7] Apache Hadoop.

<https://hadoop.apache.org/>,  
visitado el 10 de julio de 2024.

[8] Apache Airflow.

<https://airflow.apache.org/>,  
visitado el 10 de julio de 2024.

[9] Apache Superset.

<https://superset.apache.org/docs/intro/>,  
visitado el 10 de julio de 2024.