

Manual Imprescindible



Introducción a Microsoft Fabric

Guía práctica para iniciarte
en la nueva plataforma de datos de Microsoft

Diana Aguilera Reyna
Nelson López Centeno

ANAYA
MULTIMEDIA

Índice de contenidos

Cómo usar este libro.....	18
1. ¿Qué es Microsoft Fabric?.....	24
Una breve historia.....	25
Almacenamiento.....	25
Integración y gobernanza.....	27
Conclusión de la historia.....	30
Fabric según tu rol profesional.....	30
Arquitecto de datos.....	31
Ingeniero de datos.....	31
Ingeniero analítico.....	31
Científico de datos.....	32
Fabric y Power BI.....	32
Fabric y Azure.....	32
Licencias.....	33
Activar una licencia de prueba.....	35
2. Conceptos de Fabric.....	38
Inquilino.....	39
Capacidad.....	40
Dominio.....	41
Área de trabajo.....	42

Elemento	42	Punto de conexión SQL.....	96
Clasificación por cargas de trabajo.....	43	Consulta visual.....	99
Clasificación por funcionalidad.....	43	Sincronización de metadatos	100
Seguridad.....	44	Conexión desde fuera de Fabric	100
Seguridad de la plataforma.....	45	Vistas materializadas del Lakehouse.....	101
Permisos administrativos	45	¿Cuándo elegir un Lakehouse?.....	102
Controles de acceso	45	Arquitectura Medallion con Lakehouse	103
3. Navegación en Fabric	48	6. Warehouse	106
Interfaz de Fabric	49	Arquitectura del Warehouse	107
Panel de navegación.....	50	Comandos T-SQL no soportados.....	107
Área de trabajo.....	52	Operaciones más comunes con T-SQL.....	108
Configuración.....	54	Crear un Warehouse	108
Administrar acceso.....	56	Crear esquemas	109
Flujo de tareas.....	56	Crear tabla.....	109
Crear elemento	58	Insertar datos desde un archivo CSV.....	110
Carpetas	60	Consulta entre bases de datos.....	110
Filtrar elementos	61	Crear una tabla a partir de una consulta (CTAS)	111
4. OneLake: el lago de datos de Fabric	66	Crear una vista	111
Una metáfora para entender OneLake	67	Transacciones.....	112
¿Qué es un lago de datos?	68	Viaje en el tiempo.....	112
¿Qué aporta OneLake?.....	69	Instrucción T-SQL para el viaje en el tiempo	113
Accesos directos	70	Clonación de tablas.....	113
Caché de accesos directos.....	71	Puntos de restauración.....	114
Transformaciones de accesos directos	72	¿Cuándo elegir un Warehouse?	115
Reflejos (<i>mirroring</i>).....	72	7. Reflejos (<i>mirroring</i>).....	116
¿Cómo se guardan los datos estructurados en OneLake?	73	Requerimientos y limitaciones.....	117
Parquet	73	Reflejo de base de datos.....	117
Delta.....	75	Reflejo de metadatos	120
Iceberg	77	Reflejo abierto.....	122
Exploración y gestión de datos.....	78	¿Cuándo elegir un reflejo?.....	124
Catálogo de OneLake.....	79	8. Base de datos KQL	126
Explorador de archivos de OneLake.....	80	Casa de eventos.....	127
Explorador de Azure Storage.....	82	Always-on.....	128
5. Lakehouse	86	Base de datos	128
Crear un Lakehouse	88	Consultas KQL	130
Archivos y tablas.....	89	Consultas KQL desde otros elementos.....	132
Archivos	89	Replicación a OneLake.....	132
Tablas.....	90	¿Cómo acceder a las tablas Delta?.....	133
Accesos directos	91	Accesos directos	134
Transformaciones de accesos directos	94	Directivas de actualización.....	136
Transformación de archivos	94	Vistas materializadas.....	137
Transformaciones con IA	95	¿Cuándo elegir una base de datos KQL?	137

9. Otras bases de datos..... 138

SQL Database	139
Limitaciones.....	139
Crear un elemento SQL Database.....	140
Conexión desde fuera de Fabric	141
Ejemplos de consultas T-SQL.....	142
Búsquedas vectoriales.....	144
Replicación a OneLake.....	145
API para GraphQL.....	146
Panel de rendimiento	146
Copias de seguridad.....	147
¿Cuándo elegir Fabric SQL Database?.....	148
Cosmos DB.....	148
Breve introducción a Cosmos DB.....	149
Crear un elemento Cosmos DB.....	150
Ejemplos de consultas NoSQL.....	151
Conexión desde fuera de Fabric	155
Replicación a OneLake.....	155
Resumen de métricas	156
¿Cuándo elegir Fabric Cosmos DB?.....	156

10. Flujos de datos Gen2..... 158

Diferencias con los flujos de datos Gen1	159
Crear un elemento flujo de datos Gen2.....	160
Plantilla de Power Query	161
Mejoras en la interfaz de usuario	162
Pasos de consulta solo en vista previa.....	163
Almacenamiento provisional	163
<i>Staging storage</i>	164
<i>Staging Compute</i>	164
Configuración.....	165
Destinos de datos.....	165
Asignación del destino a una consulta	166
Destino predeterminado.....	170
Copia rápida	170
Requerimientos	170
Evaluador moderno.....	172
Actualización incremental.....	174
Requerimientos del origen	174
Requerimientos del destino.....	174
Configuración.....	174
Funcionamiento	175
Parámetros públicos	176
¿Cuándo elegir un flujo de datos Gen2?	177

11. Canalizaciones..... 178

Crear una canalización.....	179
Actividades.....	179
Crear una actividad	180
Configuración de una actividad	180
Ejecución manual de una canalización.....	181
Conectores de salida.....	182
Actividad de copia.....	183
Asistente para copia	184
Plantillas.....	185
Parametrización	186
Expresiones de canalización.....	186
Variables.....	187
Parámetros	189
Ejecución automática.....	191
Programación	191
Desencadenadores	191
Historial de ejecución.....	192
¿Cuándo elegir una canalización?.....	192
Orquestación.....	193
Movimiento de datos	193
Trabajo de copia	194
Crear un trabajo de copia.....	194
Asignar a destino	194
Modo de trabajo de copia	196
Revisar y guardar.....	197
Configuración posterior.....	197
¿Cuándo elegir un trabajo de copia?.....	198

12. Spark y Python

Spark en Fabric.....	201
Fabric Runtime	202
Grupos de Spark	202
Entornos	207
Python puro en Fabric.....	208
Cuadernos.....	209
Crear un cuaderno	209
Lenguajes	210
Celdas	211
Ejecución	214
Elementos de datos.....	215
Recursos	217
NotebookUtils	218
Visual Studio Code	220
Ingesta de datos	221
API REST.....	222
SQL Server	224

DataFrame.....	226	Destinos.....	274
Operaciones con DataFrame.....	226	Vista en vivo.....	277
Evaluación diferida.....	228	Panel en tiempo real.....	280
Data Wrangler.....	228	Crear un panel.....	280
Otras bibliotecas.....	230	Orígenes de datos.....	281
Carga de datos desde un Lakehouse.....	231	Mosaicos.....	282
Leer un archivo CSV.....	232	Páginas.....	283
Leer un archivo Parquet.....	233	Parámetros.....	283
Leer un archivo JSON.....	234	Actualización automática.....	284
Leer una tabla.....	235	Activador.....	285
Exploración.....	235	Funcionamiento.....	285
Mostrar el esquema.....	235	Crear un activador.....	286
Contar las filas.....	236	Obtener datos.....	287
Mostrar las primeras filas.....	236	Crear un objeto.....	288
Contar valores únicos.....	236	Crear una regla.....	289
Contar valores nulos.....	236	Centro en tiempo real.....	290
Estadísticas básicas.....	236		
Transformación.....	237	14. Elementos de ciencia de datos.....	292
Seleccionar y renombrar columnas.....	237	Predicción con la biblioteca Prophet.....	293
Columnas calculadas.....	237	Clasificación con la biblioteca MLlib.....	295
Derivadas de fechas.....	238	Ciclo de vida de un modelo con MLflow.....	297
Ejecutar un cuaderno con NotebookUtils.....	238	El elemento Experimento.....	300
Combinar dos tablas.....	238	Lista de ejecución.....	300
Dimensión Producto.....	239	Guardar como modelo de ML.....	301
Escritura hacia un Lakehouse.....	239	Crear un experimento vacío.....	303
Modos de escritura.....	240	El elemento Modelo de ML.....	303
Particionado.....	240	Aplicar esta versión.....	305
Actualización del esquema.....	241	Crear un elemento Modelo de ML con MLflow.....	306
Orquestación.....	241	Crear un modelo de ML vacío.....	307
Comando mágico %run.....	242	AutoML.....	307
Ejecutar un cuaderno con NotebookUtils.....	243	Funciones de IA.....	309
Ejecutar varios cuadernos con NotebookUtils.....	244	Funciones disponibles.....	310
Visualización.....	248	Ejemplos.....	311
Trabajos Spark.....	251	Configuración de OpenAI.....	315
Funciones de datos de usuario.....	253	Agentes de datos.....	315
Crear una UDF.....	254	Funcionamiento.....	316
Probar la función.....	256	Configuración.....	316
Conexiones.....	257		
Añadir una nueva columna.....	258	15. Gestión de la capacidad.....	320
Ejecutar la función desde un cuaderno.....	259	Operaciones.....	321
¿Cuándo elegir Spark o Python puro?.....	261	Throttling, bursting y smoothing.....	322
13. Elementos de inteligencia en tiempo real.....	262	Escalabilidad automática en Spark.....	323
Eventstream.....	264	Centro de supervisión.....	324
Crear un Eventstream.....	265	Capacity Metrics.....	326
Orígenes de datos.....	266	Chargeback.....	329
Transformaciones.....	271	FUAM.....	329

16. El modelo semántico en Fabric	332
Modos de almacenamiento	333
Crear un modelo semántico Importar o DirectQuery	334
Estrategia Fabric ETL <i>low cost</i>	334
Direct Lake.....	335
Enmarcar	335
Transcodificación	337
Temperatura de las columnas	337
Limitaciones	338
Direct Lake en OneLake.....	338
Direct Lake en SQL.....	340
¿Cuándo elegir Direct Lake?	342
Semantic Link.....	342
Utilizar la librería	343
Listar los modelos semánticos del área de trabajo.....	343
Listar las tablas del modelo semántico.....	344
Obtener los datos de una tabla	344
Relaciones del modelo	345
Medidas DAX del modelo.....	346
Evaluar una medida DAX.....	347
Comando mágico %%dax.....	349
Semantic Link Labs.....	349
17. Copilot en Fabric	352
Requerimientos de licencia y configuración.....	353
Procesamiento y almacenamiento de datos fuera de la región.....	354
Capacidades de Fabric Copilot.....	354
Copilot en el flujo de datos Gen2	355
Obtener datos	355
Otras opciones de ayuda	355
Copilot en las canalizaciones	360
Ingesta de datos	361
Transforme sus datos	362
Resumir esta canalización	363
Solucionar errores de canalizaciones	363
Copilot en el cuaderno	366
Copilot desde el panel de chat	366
Copilot desde una celda de código	370
Copilot en el Warehouse.....	372
Copilot desde el panel de chat.....	372
Copilot desde el editor de consultas.....	374
Sugerencias para usar Copilot en Warehouse	378

18. Proyectos	380
Proyecto 1: Análisis básico	381
Objetivos	381
Arquitectura	381
Flujo de trabajo.....	382
Proyecto 2: Arquitectura Medallion.....	384
Objetivos	384
Arquitectura	384
Flujo de trabajo.....	386
Proyecto 3: Ingesta con trabajos de copia y cuaderno.....	389
Objetivos	389
Arquitectura	389
Flujo de trabajo.....	390
Proyecto 4: Ingesta con canalización.....	392
Objetivos	392
Arquitectura	392
Flujo de trabajo.....	393
Proyecto 5: Transformación y modelado con cuadernos Spark.....	395
Objetivos	395
Arquitectura	395
Flujo de trabajo.....	396
Proyecto 6: Arquitectura Medallion con Warehouse en la capa Oro.....	398
Objetivos	398
Arquitectura	398
Flujo de trabajo.....	399
Proyecto 7: Almacén de datos	401
Objetivos	401
Arquitectura	401
Flujo de trabajo.....	402
Proyecto 8: ETL <i>low cost</i>	404
Objetivos	404
Arquitectura	404
Pausar y reanudar la capacidad.....	405
Transformación y modelado con Python puro.....	405
Proyecto 9: Inteligencia en tiempo real.....	407
Objetivos	407
Requerimientos	407
Índice alfabético.....	410



Cómo usar este libro

Introducción

Microsoft Fabric fue anunciada en versión preliminar en mayo de 2023 y alcanzó su disponibilidad general en noviembre del mismo año. Nosotros entendimos enseguida la importancia de esta plataforma dentro del ecosistema de datos de Microsoft, el tremendo impulso que podía dar a desarrollos actuales y las bases que sienta para el futuro, por lo que comenzamos a estudiarla y a probarla.

Hay dos aspectos que rápidamente nos llamaron la atención: la separación entre el almacenamiento y los elementos de cómputo, y la firme apuesta por formatos de datos abiertos.

En junio de 2023 fuimos al evento Power BI Days de Alicante, donde Miguel Llopis, PM de Power BI y Fabric en Microsoft, hizo una de las primeras presentaciones de la plataforma en España, si no fue la primera.

Una semana después hicimos nuestra primera presentación sobre Microsoft Fabric en el evento Power Platform Madrid, con el título "ETL con Fabric". Estaba dirigida a quienes usaban Power Query y querían conocer las nuevas posibilidades que ofrece esta plataforma.

En septiembre presentamos nuestra segunda charla sobre Microsoft Fabric, en el evento Power BI Days de Mallorca. En esta ocasión, hablamos sobre inteligencia en tiempo real, un tema que nos apasiona y sobre el que ya habíamos hecho anteriormente varias presentaciones, pero utilizando Power BI y Azure Synapse.

En febrero de 2024 nos presentamos en el evento Power BI Days de Málaga con la charla "ETL low cost con Fabric. ¿Es factible?", donde expusimos un caso de uso que aprovechaba la potencia de Microsoft Fabric para ingestar, transformar y modelar los datos, pero el modelo semántico y los informes se mantenían con licencias de Power BI. Esto permite "apagar" Fabric cuando no se esté usando y posibilita su uso con un presupuesto ajustado. De aquí salió una serie de blogs que escribimos sobre el tema.

También en febrero de 2024 presentamos una sesión introductoria a Microsoft Fabric en el evento Global Power Platform Bootcamp 2024. La charla tuvo muy buena acogida y nos motivó a iniciar una serie de vídeos en vivo en nuestro canal de YouTube que denominamos "Directos de los jueves", que se mantuvieron desde abril hasta septiembre.

Todos estos contenidos están agrupados en esta página: <https://www.dataxbi.com/fabric/>.

En enero de 2025, junto a Ana María Bisbé, lanzamos la iniciativa ADN Fabric, en la que hacemos sesiones en vivo cada dos semanas con temas de Microsoft Fabric. Por cierto, ADN son las iniciales de nuestros nombres: Ana, Diana y Nelson.

Además de estas charlas, hemos impartido varios cursos y talleres sobre Microsoft Fabric, y hemos desarrollado proyectos para nuestros clientes.

Este libro recoge parte de lo que hemos aprendido durante este tiempo, y nos hemos esforzado por hacértelo llegar de una forma clara y didáctica. Esperamos que te sea útil.

A quién está dirigido

Este libro está dirigido a profesionales que quieran tener un primer acercamiento con Microsoft Fabric, ya sea que tengan alguna experiencia con herramientas de Microsoft o no.

En algunos casos, asumimos que ya conocen Power BI, pero no es imprescindible. Nuestra intención es que, después de leer los primeros capítulos, que son introductorios, entres a Microsoft Fabric y vayas probando lo que te vamos explicando.

En Microsoft Fabric hay muchos caminos a seguir y con este libro queremos darte una guía para ayudarte a elegir el tuyo.

Qué no cubrimos en este libro

En este libro presentamos los conceptos fundamentales de Microsoft Fabric y los elementos más importantes, pero la plataforma está en constante evolución, por lo que cuando empieces a utilizarla te encontrarás novedades que no hemos cubierto en este libro. Por ejemplo, hace unas semanas se han incorporado dos nuevos elementos de inteligencia en tiempo real, los mapas y los grafos, que no hemos podido incluir.

Hay tres aspectos importantes que no hemos cubierto: la seguridad, la administración y el ciclo de vida de los proyectos. Sobre la seguridad, explicamos algunos aspectos básicos de manera general. Sobre la administración, explicamos algunas configuraciones puntuales asociadas a algún elemento y realizamos una introducción a algunas herramientas de monitoreo. En cuanto al ciclo de vida, que incluye pasar de una prueba de concepto a desarrollo, luego a *testing* y finalmente a producción, junto con el control de cambios, no lo hemos abordado en este libro.

Los tres aspectos anteriores son muy importantes en los proyectos reales y Microsoft Fabric cuenta con herramientas para implementarlos, pero en este libro introductorio preferimos no incluirlos.

Cómo está organizado

El libro tiene 18 capítulos que pueden agruparse en cuatro partes, aunque no exista esta división en el índice.

Una primera parte introductoria que abarca los capítulos del 1 al 4, donde se pone en contexto la plataforma Microsoft Fabric y se explican sus fundamentos. En el capítulo 1, se hace un recorrido rápido por la historia para fundamentar por qué consideramos que es una plataforma moderna y preparada para el futuro. En el capítulo 2, explicamos los conceptos fundamentales de la plataforma. En el capítulo 3, mostramos los elementos principales de la interfaz usuario. Y el capítulo 4 se lo dedicamos a OneLake, que consideramos que es la columna vertebral de Microsoft Fabric.

En la segunda parte, se explican los elementos individuales de la plataforma y abarca desde el capítulo 5 hasta el 14. En los capítulos 5 al 8, explicamos los elementos de almacenamiento, y del 10 al 14 explicamos los elementos que se utilizan en la ingesta, la transformación, la orquestación, el análisis en tiempo real y la ciencia de datos. Algunos capítulos son dedicados a un único elemento y en otros se agrupan varios elementos. Por ejemplo, en el capítulo 12 hemos agrupado Spark y Python; en el capítulo 13 se agrupan los elementos de inteligencia en tiempo real; y en el capítulo 14 se agrupan los elementos de ciencia de datos.

La tercera parte son los capítulos 15, 16 y 17, donde se explican diferentes aspectos. En el 15, hacemos una introducción a la gestión de la capacidad Fabric. En el capítulo 16, explicamos qué aporta Fabric al modelo semántico de Power BI. Y en el capítulo 17 mostramos cómo usar Copilot en varios de los elementos de la plataforma.

La cuarta y última parte es el capítulo 18, que contiene varios proyectos donde se utilizan los elementos de Fabric para construir soluciones. Cada proyecto tiene el diagrama de la arquitectura y referencias a los capítulos de los componentes utilizados. En algunos proyectos ofrecemos acceso a los datos e instrucciones para desarrollarlos, mientras que otros son orientativos.

Convenciones tipográficas

Como cualquier otro libro, este se encuentra dividido en capítulos que cuentan con diferentes apartados, cada uno de ellos encabezado por sus correspondientes títulos. Además, en el texto se utilizan ciertos estilos o tipografías específicas con el objetivo de hacer su lectura más rápida, facilitando la identificación de elementos como las opciones de menú, nombres de funciones, etc. Estos estilos son:

- *Cursiva*: Es un tipo que se usa para diferenciar términos anglosajones o de uso poco común. También se usa para destacar algún concepto.
- **Negrita**: Le ayudará a localizar rápidamente elementos como las combinaciones de teclas y nombres de botones. Por ejemplo, **Aceptar** para hacer referencia a un botón con ese título.
- **Monoespacio**: Diferencia elementos como los nombres de archivos y carpetas, las URL y el código (variables, funciones, etc.).
- **Menús y comandos**: Destaca sobre el texto todos los elementos relativos a interfaz de usuario, como puede ser una opción de menú, un comando o título de una ventana, etc.

También encontrará a lo largo del libro recuadros con elementos destacados sobre el texto normal, comunicándole de manera breve y rápida algún concepto relacionado con lo que está leyendo, un truco o advirtiéndole de algo. Estos recuadros serán:

NOTA:

Las notas se usan para aclarar un concepto que está utilizándose en el texto, facilitar información relacionada y tareas similares.

Archivos usados de ejemplo en el libro

En algunos capítulos mostramos ejemplos de código, hacemos algún ejercicio con archivos de datos o damos instrucciones de conexión. Los puede descargar desde la web de Anaya Multimedia (<http://www.anayamultimedia.es>). Para acceder, debemos dirigirnos al botón **Selecciona Complemento** de la ficha del libro, donde podremos descargar todos los ejemplos desarrollados para esta obra. También hemos creado un repositorio en GitHub donde se pueden descargar estos materiales: <https://github.com/dataxbi/LibroFabric>.

En el momento que se necesite el recurso, se indica dónde localizarlo.

1

¿Qué es Microsoft Fabric?

En este capítulo aprenderás a:

- Entender Microsoft Fabric a partir de la evolución de las arquitecturas de datos analíticos.
- Cómo enfocar el aprendizaje de Fabric según tu rol profesional.
- La relación de Fabric con Power BI y con Azure.
- Tipos de licencias de Fabric.
- Cómo activar una licencia de prueba.

Una breve historia

Para poder entender mejor qué es Microsoft Fabric, haremos un breve resumen de algunos conceptos sobre el almacenamiento de los datos analíticos y sobre la integración y la gobernanza.

NOTA:

En lo sucesivo, emplearemos el término "Fabric" para simplificar la referencia a Microsoft Fabric.

Almacenamiento

Comencemos con el almacenamiento y, para ello, os invitamos a un viaje en el tiempo con tres paradas, como se muestra en la figura 1.1.

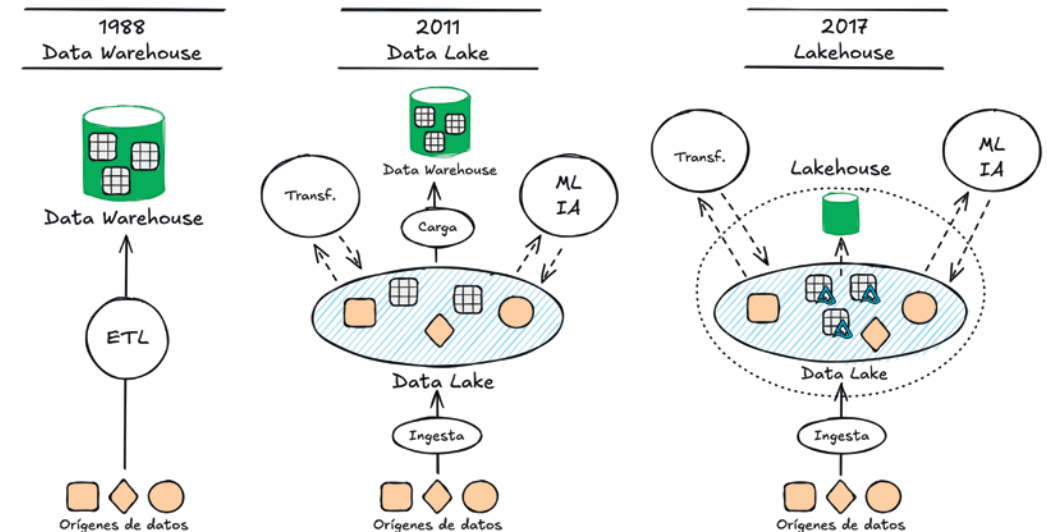


Figura 1.1. Evolución desde el Data Warehouse hasta el Lakehouse.

Data Warehouse

A finales de la década de los ochenta, aparecieron las primeras implementaciones del Data Warehouse como un repositorio donde concentrar los datos utilizados en la inteligencia de negocios y que está separado de los datos operacionales, por lo

que hay que implementar un proceso que extraiga los datos desde los orígenes, los transforme y los cargue en el Data Warehouse. Dicho proceso se conoce con el nombre de ETL por las siglas en inglés (*Extract, Transform, Load*).

En la práctica, un Data Warehouse es una base de datos SQL que sigue una estructura que se conoce como modelo dimensional, donde hay tablas de hechos y tablas de dimensiones organizadas en un esquema en estrella. Si has trabajado con Power BI, seguro que esto te suena familiar porque es el mismo principio utilizado al diseñar un modelo semántico.

Data Lake

Avanzamos en nuestro viaje y llegamos a principios de la década de 2010, cuando surge el Data Lake a partir de la aparición de los sistemas distribuidos de archivos y de las plataformas en la nube, que permiten almacenar grandes volúmenes de datos de manera segura y a bajo coste. Mientras en un Data Warehouse se almacenan datos estructurados, en un Data Lake se almacenan datos de todo tipo: estructurados, semiestructurados y no estructurados, que se ingieren desde los orígenes de datos.

Una vez que los datos están en el Data Lake, se pueden seguir transformando y volverlos a almacenar en el Data Lake con otra estructura o también se pueden cargar en un Data Warehouse o se pueden utilizar para entrenar modelos de aprendizaje automático.

Lakehouse

Terminamos este breve recorrido a finales de la misma década del 2010 con la aparición del Lakehouse, palabra que se forma a partir de unir y contraer los términos "Data Lake" y "Data Warehouse", porque aquí el Data Warehouse se transforma, separándose el motor donde se ejecutan las consultas SQL de los datos, que ahora se almacenan en un Data Lake junto al resto de los datos, lo que facilita que puedan ser utilizados por otros motores para seguir transformándolos o para entrenar modelos de aprendizaje automático.

En un Lakehouse, los datos estructurados se suelen almacenar en un formato de archivo optimizado para hacer consultas de una manera muy eficiente y, además de los datos, utiliza metadatos para tener una funcionalidad equivalente a una tabla en una base de datos SQL. Por lo general, los datos en un Lakehouse se organizan siguiendo una estructura conocida como "arquitectura Medallion", que consiste en transformar los datos por etapas: se parte de los datos sin procesar y se van refinando hasta obtener un producto final. En cada etapa se guarda una versión de los datos, lo que da lugar a distintas capas.

La explicación más habitual de esta arquitectura utiliza tres capas con nombres metálicos: Bronce, Plata y Oro. Estos nombres transmiten muy bien la idea de que los datos, como los metales, se van depurando progresivamente. Sin embargo, en la práctica puede haber más (o menos) capas y con otros nombres más acordes a las necesidades específicas del proyecto.

Integración y gobernanza

En cuanto a la integración y la gobernanza de los datos, hay tres conceptos que queremos mencionar, que mostramos en las figuras 1.2, 1.3 y 1.4.

Hub y spoke

El *hub* (figura 1.2) es un gran repositorio central donde se cargan los datos después de extraerlos desde los distintos orígenes y transformarlos para que tengan la estructura apropiada, que actúa como la "fuente única de la verdad" (*Single Source of Truth*). Los *spoke* son réplicas de pedazos de ese gran repositorio, que se utilizan para analizar un tema específico del negocio.

Los datos fluyen desde los orígenes hacia el *hub* y se distribuyen hacia los *spokes*. Hay un equipo central que es el encargado de preparar los datos y también de su gobernanza. Por ejemplo, en una gran empresa hay un Data Warehouse con todas las tablas de hechos y de dimensiones, y para los departamentos se preparan versiones que contienen solo las tablas de interés. Estas réplicas se conocen como Data Marts.

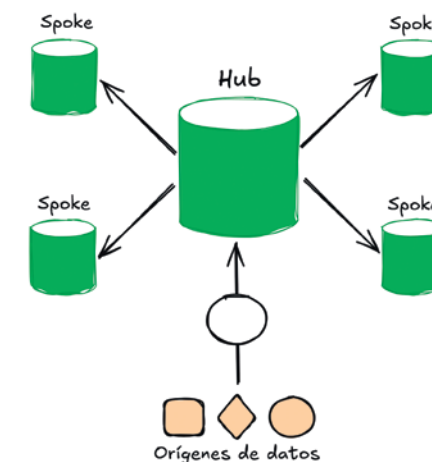


Figura 1.2. Hub y spoke.

En Power BI, un equivalente sería tener un modelo semántico centralizado que se comparte para que se creen informes a partir de él o incluso crear un modelo compuesto. También se puede implementar algo parecido utilizando flujos de datos.

Data Mesh

Data Mesh (figura 1.3) es un concepto reciente que propone descentralizar el procesamiento y la propiedad de los datos para acercarlo a los orígenes y a las personas que mejor los conocen (o los dominan), y se basa en cuatro principios:

- Responsabilidad del dominio.
- Los datos como un producto.
- Plataforma de datos de autoservicio.
- Gobernanza computacional federada.

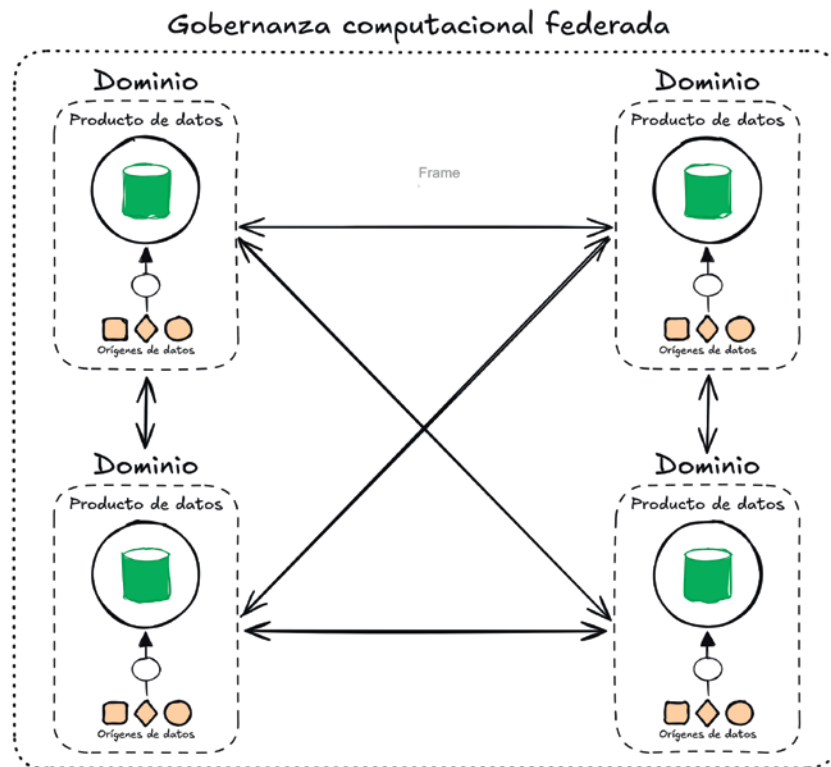


Figura 1.3. Data Mesh.

Por ejemplo, el departamento de mercadotecnia de una empresa podría considerarse un dominio que conoce muy bien sus datos y cómo analizarlos, y tiene la responsabilidad de localizar los orígenes de los datos, procesarlos y crear un producto de datos para que esté disponible para toda la empresa.

En una misma empresa pueden existir varios dominios interconectados entre sí que conforman una malla; de esta metáfora surge el término "Mesh".

El producto de datos podría ser un modelo semántico y varios informes de Power BI. Pero también podría ser una API o un modelo de aprendizaje automático o un asistente conversacional al estilo ChatGPT.

A nivel de empresa, debe existir una plataforma que facilite a las personas de los dominios conectarse a los orígenes de datos, procesarlos y crear los productos de datos, pero que al mismo tiempo permita establecer las reglas de gobernanza y facilitar el descubrimiento de los productos de datos.

Data Fabric

Data Fabric (figura 1.4) es también un concepto reciente que, en esencia, propone simplificar y democratizar el acceso a los datos a partir de agregar capas de servicio que interconecten los distintos orígenes y procesos de datos y analice sus metadatos de manera automatizada con la ayuda del aprendizaje automático y la inteligencia artificial.

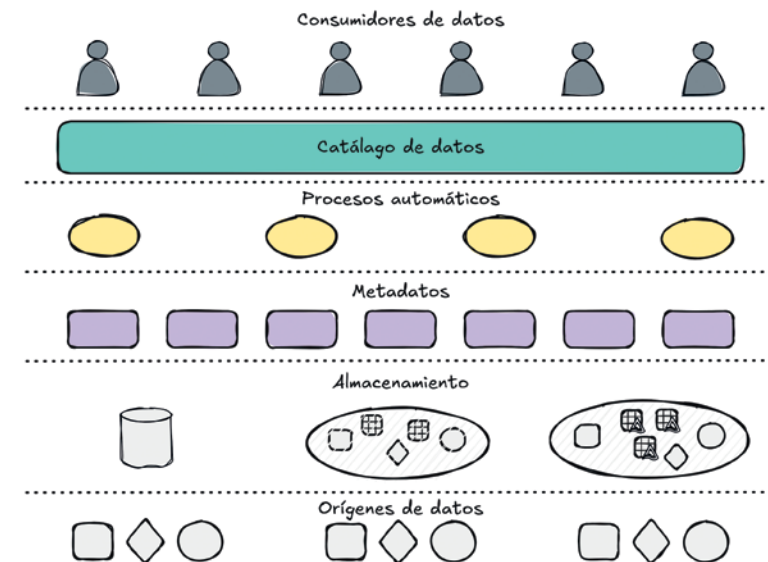


Figura 1.4. Data Fabric.

Imagina un edredón: varias capas cosidas que trabajan al unísono para ofrecer calidez y confort. Es a ese tejido multicapa al que hace referencia el nombre "Fabric".

Entre los resultados de este proceso automatizado, estarían un catálogo de datos aumentado, un grafo de conocimiento, un motor de recomendaciones o incluso la creación automática de canalizaciones de datos.

En la figura 1.4 estamos representando un proceso que va desde la parte inferior con los orígenes de datos hasta la parte superior con los consumidores de datos y, entre ellos, está la capa de almacenamiento y, sobre estas, varias capas adicionales que corresponderían a Data Fabric.

Conclusión de la historia

Podemos extraer varias conclusiones de los párrafos anteriores:

- Hay una tendencia a separar el motor analítico, por ejemplo, el motor que hace las consultas SQL, de los datos.
- Lo anterior es posible porque hay nuevos formatos de archivos que permiten almacenar los datos estructurados de forma óptima y que además son de código abierto.
- El producto final del análisis de datos ya no es solo un modelo dimensional, sino que hay otros requerimientos, como, por ejemplo, preparar los datos para un agente conversacional.
- Hay varias formas de organizar los datos en una empresa, que van desde tenerlo todo centralizado a dar autonomía a los departamentos para que creen sus productos de datos, pero también se puede llegar a un punto intermedio.

¿Y qué relación tiene esta historia con Fabric? Pues bien, Fabric es una plataforma que nos proporciona herramientas para implementar muchos de los escenarios que mencionamos antes, de forma relativamente sencilla y a través de una interfaz muy similar a la de Power BI.

Fabric según tu rol profesional

Las herramientas de Fabric son muy diversas, porque es una plataforma flexible capaz de adaptarse a múltiples escenarios. Pero esa misma amplitud puede resultar abrumadora al inicio, con tantas opciones que a veces no sabemos por dónde empezar, y corremos el riesgo de caer en la "parálisis por análisis".

Para evitarlo, te recomendamos que no intentes dominar todo de golpe. En su lugar, céntrate en las herramientas que necesites para el rol que desempeñas (o para el que aspiras). Ten en cuenta que muchas tareas se pueden resolver con distintas herramientas, así que elige la que mejor se adapte a tu proyecto. También verás que una misma herramienta puede servir en varios contextos.

Para que te sirva de referencia, a continuación, enumeramos los roles más comunes en proyectos de analítica de datos y analizamos su relación con Fabric. Esta lista no es exhaustiva y excluimos roles de aspectos que no se abarcan el libro, como por ejemplo los especialistas en gobernanza o en visualizaciones.

Arquitecto de datos

El arquitecto de datos es el responsable de diseñar la estructura global del proyecto y de decidir las herramientas a utilizar. Por lo tanto, necesita un conocimiento general de Fabric, saber para qué utilizar cada herramienta y qué elementos hay que tener en cuenta a la hora de escoger una herramienta entre las opciones disponibles.

Ingeniero de datos

El ingeniero de datos es el que construye y mantiene los procesos que traen los datos desde sus orígenes, los limpia y los transforma. Tiene que garantizar la robustez de estos procesos para minimizar los errores y, en caso de que ocurran, generar registros y alertas.

En Fabric encontrará herramientas propias de Microsoft, que le serán familiares si conoce Microsoft Azure, pero también podrá trabajar con Python, Spark o con SQL.

Ingeniero analítico

El ingeniero analítico es un híbrido entre el analista de negocio y el ingeniero de datos, y es un rol muy habitual en los proyectos con Power BI donde utiliza Power Query para transformar los datos y construir el modelo dimensional; también tiene conocimientos de DAX para crear las medidas.

Este rol se sentirá a gusto en Fabric porque tiene las mismas herramientas, pero también tiene una versión mejorada de los flujos de datos y más opciones para trabajar con los modelos semánticos en el servicio y seguro que se animará a utilizar otras herramientas.

Científico de datos

El científico de datos utiliza el aprendizaje automático y la inteligencia artificial para hacer análisis predictivos y prescriptivos y su principal herramienta de trabajo son los cuadernos (*Notebooks*) con Python o Spark.

En Fabric tiene a su disposición esos mismos cuadernos, así como muchas de las librerías Python que necesita y con la posibilidad de instalar otras. También encontrará un entorno basado en MLflow, donde puede definir los experimentos y manejar todo el ciclo de los modelos, que a su vez podrán ser utilizados por otras herramientas.

Fabric y Power BI

Power BI es uno de los componentes de la Power Platform, junto a otras herramientas enfocadas al usuario de negocios, como Power Apps o Power Automate. Y ahora Power BI también es uno de los componentes de Fabric (figura 1.5).

Esta dualidad nos parece bien, porque entendemos que hay muchos proyectos donde se va a seguir utilizando Power BI sin Fabric, como también habrá proyectos de Fabric que no utilizarán Power BI.

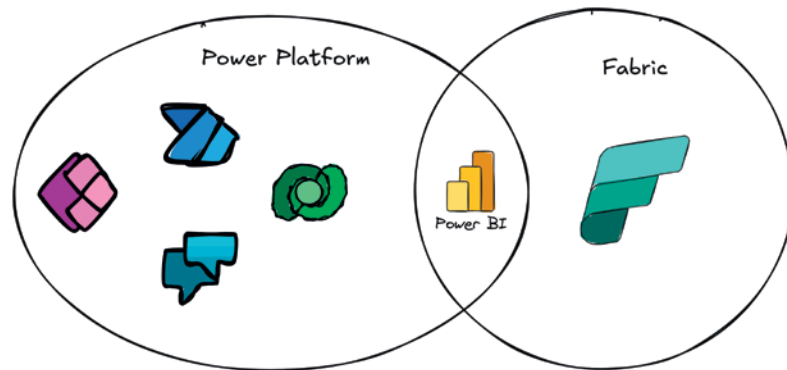


Figura 1.5. Fabric y Power BI.

Fabric y Azure

Fabric está alojado en Azure, que es la nube de Microsoft, y hace uso de sus servicios de conectividad, seguridad y de almacenamiento. Algunas de las herramientas de Fabric han evolucionado desde servicios que existen en Azure o utilizan directamente algunos servicios de Azure.

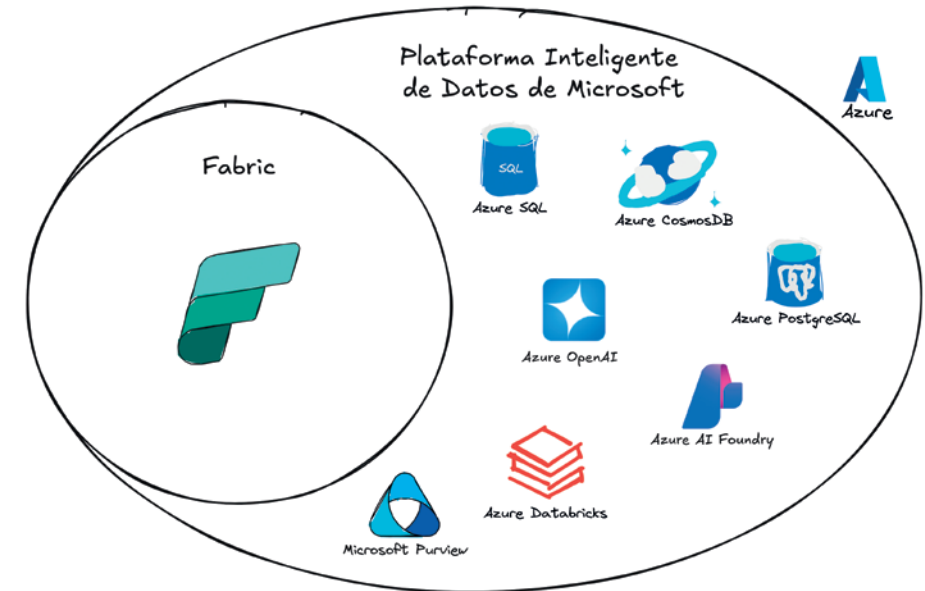


Figura 1.6. Fabric y Azure.

Además, Fabric es parte de algo más grande llamado Plataforma Inteligente de Datos de Microsoft, donde se agrupan todas los productos y servicios de Azure para el trabajo con los datos, por ejemplo, como se muestra en la figura 1.6: bases de datos como Azure SQL o Cosmos DB o Azure PostgreSQL, servicios de IA como Azure OpenAI o Azure Foundry, el servicio de análisis de datos Azure Databricks y el servicio de gobernanza de datos Microsoft Purview.

Al ser parte de la misma plataforma, se logra un gran nivel de integración entre Fabric y estos servicios.

Licencias

Las licencias de Fabric están asociadas a una capacidad, que es el nombre que reciben los recursos de procesamiento y memoria disponibles para que las herramientas ejecuten sus tareas.

Si vienes de Power BI, las licencias de Fabric son similares a la licencia *Premium* por capacidad que existía en Power BI. De hecho, ya no es posible adquirir nuevas licencias Power BI Premium por capacidad y las empresas que tienen contratos vigentes deben migrar a una licencia Fabric cuando se venza el contrato.

Hay tres modalidades de licencias de Fabric:

- Licencia de prueba.
- Pago por uso.
- Reserva.

La **licencia de prueba** es gratuita y la podemos utilizar durante 60 días para aprender o para hacer pruebas de concepto o simplemente para probar lo que ofrece Fabric. Esta licencia se puede activar desde el servicio de Power BI, y en el próximo apartado explicamos cómo hacerlo.

La **licencia de pago por uso** se compra en Azure y se puede cancelar en cualquier momento porque no hay ningún compromiso de permanencia. Además, tenemos la posibilidad de pausarla cuando no la necesitemos y durante el tiempo que esté en pausa no se cobra.

La modalidad de **reserva** también se compra en Azure. Tiene un compromiso de permanencia de un año y a cambio se aplica un descuento de un 40 % con relación al precio del pago por uso. Con esta modalidad los costes son similares a los de Power BI Premium por capacidad.

Hay varios niveles de capacidad y por tanto de precio: F2, F4, F8, F16 y así hasta la F2048. Para dar una idea, una capacidad Fabric F64 es equivalente en rendimiento y en precio a una capacidad Power BI Premium P1. La F2 es una capacidad con unos recursos muy modestos, pero a un precio bastante razonable, que se puede utilizar en proyectos pequeños. La licencia de prueba tiene una capacidad equivalente a una F64.

En la figura 1.7 te mostramos los distintos niveles (SKU) con su equivalente en Power BI y te damos una idea de los precios, primero el precio mensual con reserva, luego el precio mensual en pago por uso si no se pausa en ningún momento y en la última columna el precio por hora en pago por uso, aunque en realidad el metraje se hace por segundo.

Los precios están en euros y son solo para referencia, porque varían por región geográfica y también dependen de la tasa de cambio. Los precios actualizados los puedes ver en esta página: <https://azure.microsoft.com/es-es/pricing/details/microsoft-fabric/>. El precio por hora lo mostramos porque hay escenarios donde se puede utilizar una capacidad solo durante un periodo de tiempo, por ejemplo, para hacer una ETL y luego pausarla. O también se podría tener un nivel de capacidad, digamos una F64, y en determinados periodos de tiempo aumentar el nivel a una F128 porque hay un pico en la carga. Pero estos son escenarios avanzados, que por ahora requieren hacer un desarrollo propio para poder manejar la capacidad de una forma dinámica.

SKU Fabric	SKU Power BI	Precio Mes con Reserva	Precio Mes Pago por uso	Precio Hora Pago por uso
F2		€ 183	€ 308	€ 0,42
F4		€ 366	€ 616	€ 0,85
F8	EM/A1	€ 733	€ 1.232	€ 1,69
F16	EM2/A2	€ 1.466	€ 2.465	€ 3,38
F32	EM3/A3	€ 2.932	€ 4.930	€ 6,76
F64	P1/A4	€ 5.864	€ 9.861	€ 13,51
F128	P2/A5	€ 11.728	€ 19.723	€ 27,02
F256	P3/A6	€ 23.457	€ 39.447	€ 54,04
F512	P4/A7	€ 46.914	€ 78.894	€ 108,07
F1024	P5/A8	€ 93.828	€ 157.788	€ 216,15
F2048		€ 187.655	€ 315.576	€ 432,30

Figura 1.7. Precios de Fabric.

El **almacenamiento** en Fabric **se cobra aparte**, aunque es muy barato. Para dar una idea, el precio actual en euros en la región oeste de Europa es de 0,0204 euros por GB al mes.

Hay que tener en cuenta, que para **publicar contenido Power BI** hay que tener una licencia Power BI Pro.

Y para **ver contenido Power BI** también se necesita licencia Power BI Pro para las capacidades F2 a F32. A partir de F64 se puede acceder al contenido Power BI con una licencia gratuita (que es diferente a la licencia de prueba).

Activar una licencia de prueba

Para activar la licencia de prueba de Fabric, debemos tener alguna licencia de usuario de Power BI. La manera más sencilla es entrar al servicio de Power BI y abrir el administrador de cuenta haciendo clic sobre el icono con nuestro avatar que está en la esquina superior derecha, donde veremos un botón que nos permitirá iniciar la prueba. En la figura 1.8 se muestra el administrador de cuenta con el botón y la secuencia de pasos.

Si no se muestra el botón de activación, otra opción para activar la licencia de prueba es entrar en Mi área de trabajo e intentar crear un nuevo elemento del tipo Flujo de datos Gen2, que es una de las herramientas de Fabric, y en ese momento se activa la licencia de prueba.

Si esto tampoco funciona, puede ser que en tu empresa no esté habilitada Fabric o que no dejen crear licencias de prueba o que hayan llegado al límite de las licencias de prueba que se permiten. Te sugerimos que contactes a los

responsables de la administración de Power BI o que le preguntes a tus colegas si alguno tiene licencia de prueba para que te cree un área de trabajo que tenga asignada su licencia gratuita y que te dé acceso a dicha área de trabajo.

Si no logras tener una licencia gratuita en tu empresa, puedes intentar crear una cuenta de Fabric para ti registrándote en la página <https://app.fabric.microsoft.com> con una dirección de correo de tu propiedad. Ten en cuenta que dicha dirección de correo tiene que pertenecer a un dominio propio y no puede ser de un servicio público de correo como live.com, gmail.com o similares. Tienes que poder recibir correos en dicha dirección porque te enviarán un código de comprobación.

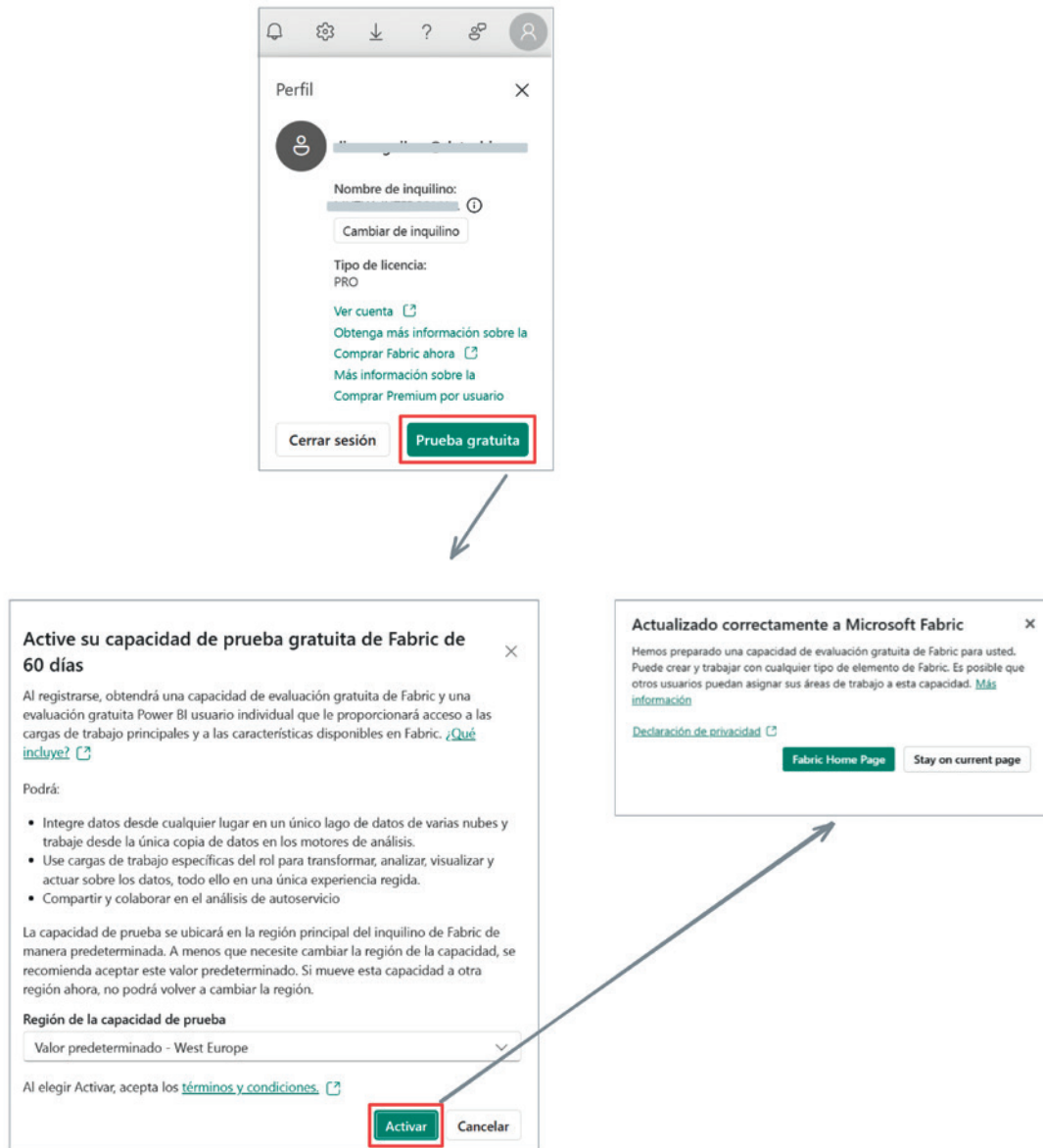


Figura 1.8. Activar licencia de prueba.

Introducción a Microsoft Fabric



Manual Imprescindible

Descubre Microsoft Fabric, la nueva plataforma de datos de Microsoft que amplía las capacidades de Power BI y permite implementar procesos de datos modernos de forma sencilla y unificada.

Basada en los mismos principios que hicieron exitoso a Power BI, Microsoft Fabric incorpora capacidades avanzadas de datos en un entorno completamente integrado, preparado para aprovechar todo el potencial de la inteligencia artificial generativa.

Este libro te guía paso a paso, desde los conceptos clave y su interfaz de usuario hasta los principales elementos de la plataforma: almacenamiento, ingesta, transformación, orquestación, análisis en tiempo real y ciencia de datos que te ayudarán a dominar Microsoft Fabric.

Tanto si vienes de Power BI como si te inicias en el mundo del análisis de datos, este libro es para ti.