

Capítulo 1: Fundamentos en Preparación de Datos Inteligente

¿Qué es Power Query?



Power Query Logo

🔗 NOTA: Dataflows

Un *flujo de datos* o *Dataflow* es una solución ETL alojada en la nube creada con *Power Query Online*, la cuál es posible implementar en múltiples proyectos y plataformas, con fácil acceso para los usuarios desde cualquier parte.

Para más información:

<https://www.youtube.com/watch?v=bnZEM1zupv4&t=488s>

Power Query es un motor de preparación de datos el cual permite: *adecuar*, *limpiar*, *formatear*, *consolidar*, *crear*, *refinar* y hasta *resumir* datos de forma veloz, intuitiva y coherente sin apenas código.

Power Query permite reutilizar las conexiones y la lógica de preparación a lo largo de múltiples proyectos (*utilizando parámetros*) y plataformas (*utilizando flujos de datos*). Ejecutando los ajustes requeridos en el proyecto concreto, todo ello sin ninguna experiencia de programación.

Versión de Escritorio y Online

Existen dos grandes versiones de *Power Query*:

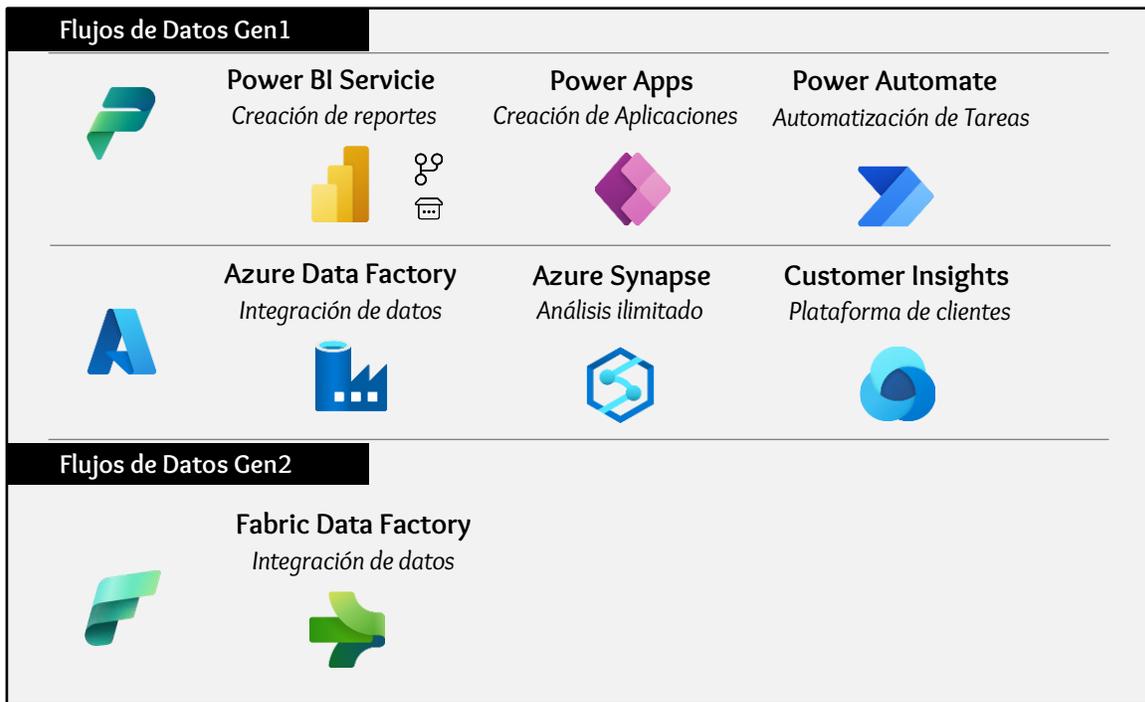
- Power Query de Escritorio (Desktop)
- Power Query en la Nube (Online) – Flujos de datos

Estas dos versiones permiten que la tecnología de preparación de datos se encuentre en múltiples aplicaciones y servicios de *Microsoft*, y crece con el transcurrir del tiempo

— POWER QUERY DE ESCRITORIO SE ENCUENTRA EN:



— POWER QUERY EN LA NUBE SE ENCUENTRA EN:



Lenguaje de Programación M

Las transformaciones de *Power Query* son expresadas internamente en un lenguaje de programación denominado: Lenguaje M, resulta que: **La interfaz es sólo la punta del iceberg.**

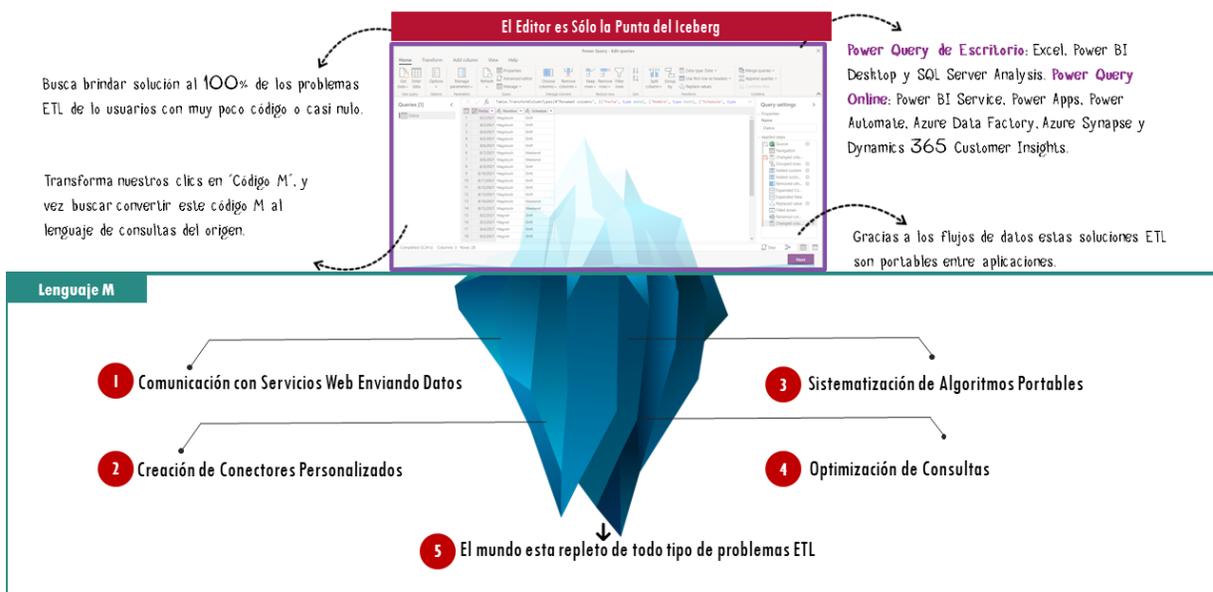


Figura 1. 1– Analogía del Editor de Power Query y el Lenguaje M con un Iceberg

Conceptos Fundamentales

NOTA

Como se señala en el párrafo de introducción de la derecha, el presente capítulo enseñará diversidad de conceptos y definiciones, en otras palabras, tiene bastante teoría, empero, queremos hacer hincapié en que la narrativa del resto del texto está ligada a la práctica inmediata, estudiando transformaciones que ocurren en los escenarios laborales.

Para conseguir el mejor resultado al emprender la preparación de cualquier receta de un libro de cocina ***es primordial conocer los ingredientes que la compondrán***, los utensilios implicados en la ***cocción y tener una imagen clara del plato que deseamos conseguir***. Por esta razón estimado(a) lector(a), antes de empezar a “cocinar distintas recetas con Power Query”, primero, haremos un recorrido por ciertos conceptos claves que serán una guía para sacar el máximo provecho a esta potente herramienta; puede que al principio no nos parezcan familiares e incluso suenen extraños, pero no debes preocuparte, ya que al ir practicando se irán afianzando en tu interior, sin embargo, es crucial tener estos ingredientes listos en la cocina.

Entendimiento de los Datos

Preparación de Datos

La *preparación de datos* consiste en una primera instancia en el ***entendimiento de los datos***, y este a su vez se componen de cuatro pilares:

- a. Exploración
- b. Descripción
- c. Selección
- d. Calidad

Proceso ETL

En segunda instancia se compone del ***proceso ETL*** (*extracción, transformación y carga*), así:

$$f(x)$$

Preparación de Datos =
Entendimiento de Datos + Proceso ETL

Estas dos partes del algoritmo eran partes sucesivas entre sí, dicho de otro modo, primero realizábamos el *entendimiento de los datos* y luego el *proceso ETL*, a pesar de ello, estos conceptos van mutando con el tiempo, principalmente por los esfuerzos de *Microsoft* de hacernos la vida más fácil, por lo previo, la parte de entendimiento se puede realizar con *Power Query* una vez extraído (*cargados*) los datos, todo gracias a la funcionalidad de perfil de datos (*Data profiling*) que nos facilita esta empresa.



Entendimiento de Datos

- **Exploración:** Consiste en el acceso y navegación a los datos pedidos en el requerimiento inicial, con ellos se busca auditar si existen los datos completos o parciales.
- **Descripción de los Datos:** Trata de entrar a conocer a profundidad que contienen las tablas de la base de datos según el negocio.
- **Selección de Datos:** Consiste en escoger las tablas pertinentes para el proceso de inteligencia de negocios, implicadas en las preguntas de negocio que se deben responder.
- **Calidad de Datos:** Verificar la calidad haciendo preguntas como: *¿Están completos? ¿Contiene errores? ¿Están en formato tabular? ¿En cuál tipo de forma están? ¿Existen valores faltantes? ¿Es una estructura ideal desde el punto de vista de modelado de datos? ¿Su lectura es detectada de forma correcta?*



Exploración



Descripción



Selección



Calidad

Integración de Datos

El Proceso ETL/ELT

El proceso ETL hace parte de la integración de datos, en otras palabras, es una pieza importante en la orquestación y automatización del movimiento y transformación de datos estructurados y no estructurados, las letras ETL significan:

- a. Extracción
- b. Transformación
- c. Carga



Las siglas derivan del inglés: **Extract**, **Transform** and **Load**. Este proceso consiste en llevar datos de múltiples orígenes, para luego hacer la limpieza necesaria, y finalmente dejarlos en el destino listos para el consumo, con un contexto que los dota del significado adecuados para proporcionar información significativa a los analistas, científicos de datos y responsables de decisiones empresariales.

— Históricamente el proceso ha lucido así:



Figura 1. 2– Proceso ETL (Extracción, Transformación y Carga) Histórico

Los datos son extraídos de sistemas de almacenamiento relacionales, no relaciones y de otras fuentes para luego pasar a un área exclusiva de transformación, esto incluye limpieza y optimización para el análisis, una vez culminado, son cargados a una base de datos analítica u OLAP (*Online analytical processing*).

Sin embargo, *el proceso ETL también ha ido evolucionando gracias a los avances en las tecnologías, llegando así al proceso ELT*, permitiendo que la etapa de transformación y análisis se pueden realizar en un mismo motor analítico.

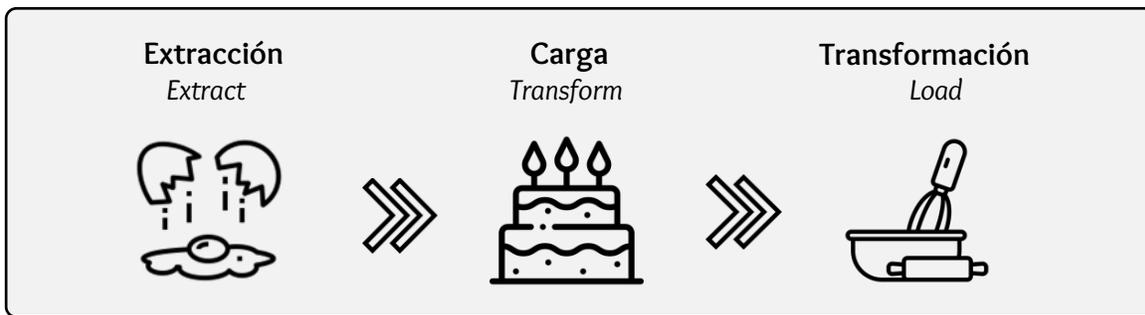


Figura 1. 3 – Proceso ELT (Extracción, Carga y Transformación)

La ventaja de la modernización del proceso ETL recae en que el proceso de inteligencia de negocios se hace en un mismo lugar, esto estriba en mayor control de la ejecución.

Pausemos un Momento



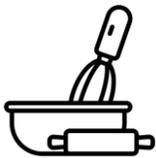
Aunque la preparación de datos es la suma del entendimiento y el proceso ETL, ciertamente en este libro pasaremos gran parte del tiempo en la fase de transformación del proceso ETL, esto se debe a que allí surgen la gran mayoría de desafíos y casos de transformación que van en aumento, esto no quiere decir que la etapa de extracción no se contemple y que también cuente con sus retos: *Véase el problema de lectura incompleta de conector Excel más adelante*

A continuación, presentamos una definición de cada etapa del proceso ETL:



Extracción

- **Extracción (Extract):** Aprovechando la flexibilidad de *Power Query* podemos conectarnos y extraer datos desde distintos orígenes como, por ejemplo: *Archivos de Texto*, *Archivos de Excel*, *Carpetas* o diversos *Gestores de Bases de Datos locales*, o en la nube, *APIs*, etc. Lo más interesante es que nunca ponemos en riesgo la integridad de los datos y podemos manipularlos en el Editor de *Power Query* a nuestro antojo.



Transformación

- **Transformación (Transform):** Después de conectarte al origen y extraer los datos, si es necesario, tienes que limpiarlos, consolidarlos o incluso enriquecerlos para alcanzar un *Formato Tabular* y/o reestructurarlos para un diseño apropiado de modelado, todo esto utilizando diversas técnicas de limpieza y consolidación de datos.



Carga

- **Carga (Load):** Con los datos en formato tabular y modelado para responder las preguntas de negocio, es momento de cargarlos en algún destino, si estamos en *Excel* puede ser en una *hoja del libro* o crear sólo la *conexión*; si estamos en *Power BI de Escritorio* podemos cargarlos para la utilización del *DAX Engine*, si estamos *Azure Data Factory* se puede publicar en almacenes de datos como en *Azure Synapse Analytics*, para que los consuman aplicaciones de inteligencia empresarial, o si estamos en *Power BI Service* dejarlo disponible en *Dataverse* y así sean consumidos por los servicios del *Power Platform*, o con un *datamart* para una base de datos de autoservicio.

— El proceso ETL debe cumplir con ciertos puntos críticos:

- **Soporte para captura de datos cambiantes:** Actualización sencilla y programada sin necesidad de hacer una carga completa de los datos nuevamente, además, propiciar sólo la actualización de los datos recientes de tal forma de garantizar rapidez.
- **Auditoria:** capacidad para poder realizar un proceso de revisión rápida y certera.
- **Manejo de múltiples formatos:** Para manejar datos de múltiples orígenes, es necesario tener herramientas para llevarlos a un formato estándar.
- **Tolerancia a errores:** En cualquier sistema ocurren errores, un buen proceso ETL debe ser robusto para una recuperación ágil.
- **Velocidad:** El sistema debe ser capaz de manejar macrodatos.

NOTA

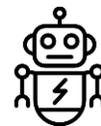
Algunas funcionalidades que ayudan a *Power Query* a cumplir los puntos críticos del proceso ETL son:

- Actualización incremental
- Diagnóstico de sesión, consultas y pasos
- Métodos de transformación de datos y Lenguaje M
- Técnicas para manejo de excepciones
- Plegado de consultas

Consulta o Query: La Esencia de Power Query

Power Query no lucha solo en la cruzada por convertir datos en tablas que pueden ser analizadas, bajo su mando cuenta con las consultas; componente sustancial en el universo del Proceso ETL.

Una consulta está formada por una serie de pasos creados en un orden específico definidos en el propio lenguaje de expre-



¿Qué es una
Consulta?
(Query)

... siones de *Power Query*, denominado *Lenguaje M*, dichos pasos tienen como objetivo realizar el proceso ETL para posterior consumo y uso en las soluciones creadas, por ejemplo: en *Excel*, *Power BI* o *Azure Synapse Analytics*. En términos simplificados:


Bloque de Trabajo
 (Query)

Una Consulta es un Bloque de Trabajo

Para contemplar el poder de una consulta y no malinterpretar su lugar dentro de *Power Query*, examinemos que hace desde distintas perspectivas:

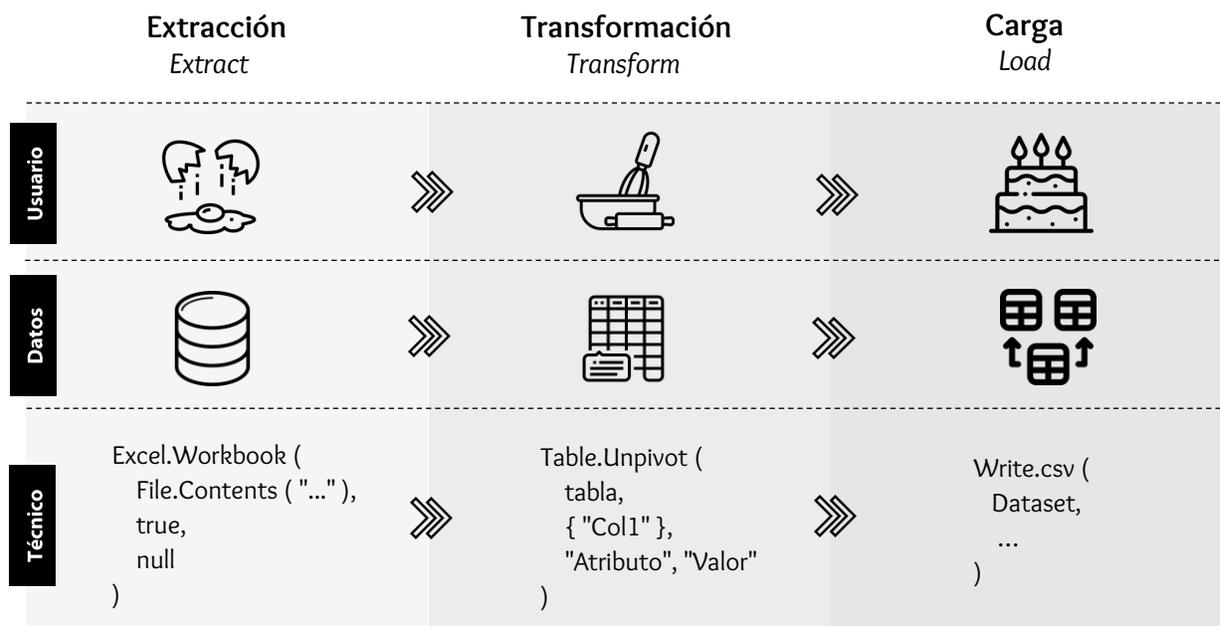


Figura 1. 4 – Una Consulta Desde Distintas Perspectivas

- **Usuario:** Nos permite conectarnos a distintos orígenes, extraer sus datos, transfórmalos si es necesario, para luego cargarlos en *Excel*, *Power BI* u Otro.
- **Datos:** Una tabla pasa por las tres etapas del proceso ETL, convirtiendo datos o una tabla sin formato alguno en una con *Formato Tabular*.
- **Técnico:** Una consulta son líneas del *Lenguaje M*, que ejecuta el Proceso ETL.

Objetivo: Formato Tabular y Modelado

Esta serie de conceptos que te hemos venido presentando como parte esencial del funcionamiento de *Power Query* tiene un gran objetivo común, y es: **aplicar el formato Tabular a todas las tablas**, para ello cada una debe cumplir con estos requisitos:

- Columnas con etiquetas adecuadas
 - ✓ Minimizar espacios en blanco
 - ✓ Minimizar caracteres no imprimibles
 - ✓ Minimizar acentos y tildes
 - ✓ Primera letra en mayúscula (Pascal Case)
- Sin totales ni grandes totales
 - ✓ Sin subtotales
 - ✓ Sin totales generales
 - ✓ Sin grandes totales
- Sin agrupación a nivel de filas (rótulos)
- Sin agrupaciones a nivel de columnas (niveles)
- Sin columnas no apropiadas
 - ✓ Sin columnas de error
 - ✓ Sin columnas vacías
 - ✓ Sin columnas no útiles (no congruentes)
 - ✓ Sin columnas duplicadas
- Sin filas no apropiadas
 - ✓ Sin filas de error
 - ✓ Sin filas vacías
 - ✓ Sin filas no útiles (no congruentes)
 - ✓ Sin encabezadas como primera fila

Formato Tabular

NOTA

NOTA 1

El formato tabular consiste en tener todos los datos de tal forma que las filas describan un pedacito de información concreta, optimizada y ordenada para una manipulación amigable y un análisis efectivo; este formato se compone por columnas o campos, y en filas registros, conformando así el formato clásico de una tabla o como están estructurados en un gestor de bases de datos.

NOTA 2

Pascal Case es una convención en programación para los nombres, de tal forma que la primera letra de cada palabra sea en mayúscula y las demás en minúsculas.

★ **NOTA**

Modelado

Sabiendo que el modelo de datos es el conjunto de tablas, relaciones, etc.

Modelar datos: consiste en crear u organizar hechos y dimensiones de tal forma que estén orientadas a responder a las preguntas de negocio planteadas inicialmente, esto para abrazar un buen rendimiento y facilitar los procesos posteriores, cómo la parte analítica con la creación de indicadores. Esto siempre va encaminado a un modelo en estrella.

El presente libro no es un libro sobre modelado.

- Columnas con tipos de datos apropiados
- Para elementos de columnas
 - ✓ Elementos estándares
 - ✓ Elementos bien escritos
 - ✓ Eliminar espacios innecesarios
 - Espacios dobles o excesivos
 - Espacios al principio y al final
- Eliminar caracteres no imprimibles
 - Tabulación
 - Espacios especiales
 - Enter
 - Etc.
- Sólo columnas de interés para el análisis
- Sólo registros de interés para el análisis

A continuación, te presentamos un ejemplo sencillo, pero claro de una tabla con el formato tabular óptimo:

| Producto | Unidades | Precio | Costo | Cantidad | Categoría |
|-------------|----------|--------|-------|----------|-----------|
| Pasta | 40 | 6 | 4 | 150 | Alimentos |
| Spaguetti | 35 | 5 | 9 | 237 | Alimentos |
| Cereal | 46 | 3 | 3 | 156 | Alimentos |
| Frijol | 50 | 9 | 8 | 291 | Alimentos |
| Mantequilla | 60 | 8 | 8 | 174 | Alimentos |
| Queso | 39 | 7 | 7 | 269 | Alimentos |
| Chocolate | 37 | 4 | 9 | 266 | Alimentos |
| Café | 33 | 9 | 2 | 199 | Alimentos |
| Frijol | 50 | 9 | 8 | 291 | Alimentos |
| Mantequilla | 60 | 8 | 8 | 174 | Alimentos |
| Queso | 39 | 7 | 7 | 269 | Alimentos |

Figura 1. 5 – Formato Tabular Óptimo

Tener a la mano esta lista es una opción buena, por ello dejamos de manera externa su descarga a modo de check list.

¡Descargar!



Para cualquier consulta de la lista de criterios a cumplir para el formato tabular se puede descargar el siguiente check list: <http://bit.ly/2X7y3IQ>

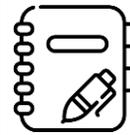


– Definición:

Formato Tabular o Tabular Format



Es un formato que presenta todos los datos en arreglos de filas (registros) y columnas (campos), sin incluir ninguna especie de etiquetas ni totales a parte de los nombres de los campos. Se dice que es óptimo cuando cumple todo el check list.



– Esquema visual del formato tabular o tabular table.

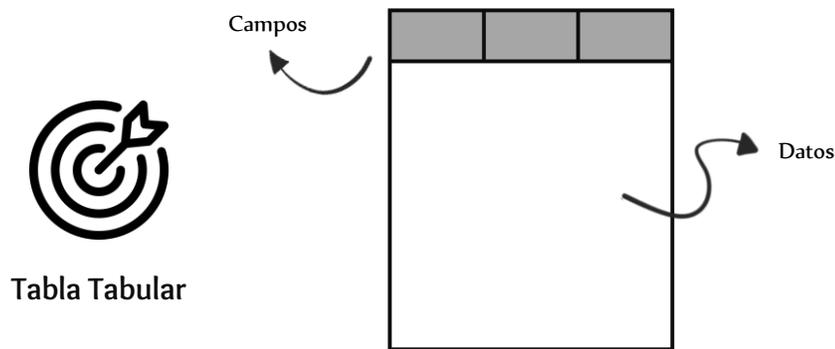


Figura 1. 6 – Esquema del Formato Tabular

Pausemos un Momento



Los esquemas visuales de representación de formatos de tablas son grafos que tienen como intención percibir la esencia de la estructura en la cual están organizados los datos, estos aparecen a lo largo de todo el libro y tiene como objetivo servir como hoja de ruta para una identificación ágil, en el artículo *Topología de Tablas es Power Skill Tech* recopilamos de estos esquemas y cómo solucionarlos.

El Editor de Power Query

NOTA

Una maravilla para destacar de crear una solución a un problema ETL utilizando *Power Query Online* o *Dataflow* en un entorno *Power Platform* o *Fabric*, es que esta solución puede ser consumida fácilmente en: *Power Apps*, *Power Automate*, *Power BI (Desktop y Service)* y hasta *Excel* por múltiples usuarios y proyectos, siempre que estos cuenten con los permisos.

En otras palabras, crear y mantener una vez, pero utilizar en muchas partes.

Power Query cuenta con su propia ventana de trabajo, es como si fuera una aplicación en sí misma, sin embargo, está “entrelazada” con *Power BI*, *Excel* o la aplicación en cuestión; o el servicio particular si se trata de la nube (*más detalles adelante*).

¿Cómo ir al Editor de *Power Query Desktop*?

Existen dos caminos bien definidos para ir al *Editor de Power Query* en escritorio, la primera es ir a secas, para ello:

1. Abrimos un archivo en blanco de *Power BI*.
2. Vamos a la pestaña *Inicio*, grupo *Datos externos*, desplegamos las opciones de *Editar consultas* y pulsamos clic en *Editar Consultas*:

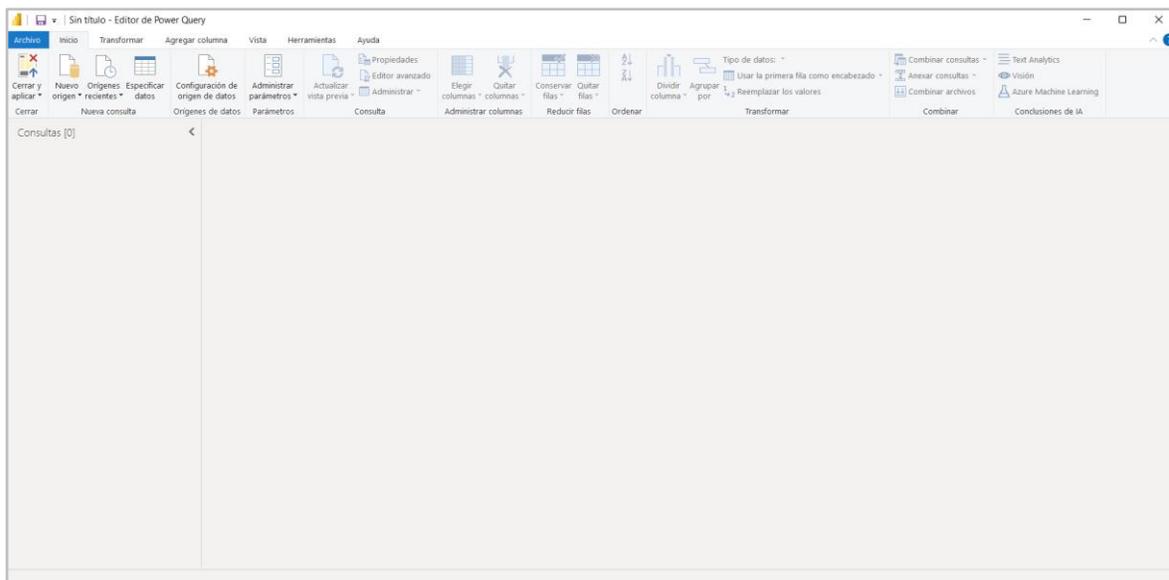


Figura 1.7 – Ventana del Editor de Power Query en Power BI

Otra alternativa es cargar datos, por ejemplo, si tomamos la tabla de *Excel* en el siguiente archivo:

¡Let's Play!



ARCHIVO DE TRABAJO: *En el compendio de archivos y carpeta Capítulo 1, puedes encontrar el archivo de Excel con nombre: CAPIEJI-SKUProductos.xlsx.*



— Ahora:

1. Abrimos un archivo en blanco de *Power BI*
2. *Inicio* → *Datos Externos* → *Obtener datos* → *Excel*
3. Navegamos en nuestro computador hasta encontrar el archivo de Excel: *CAPIEJI-SKUProductos.xlsx*
4. En el cuadro de diálogo seleccionamos el elemento *SKUProductos* y pulsamos clic en el botón *Editar*.

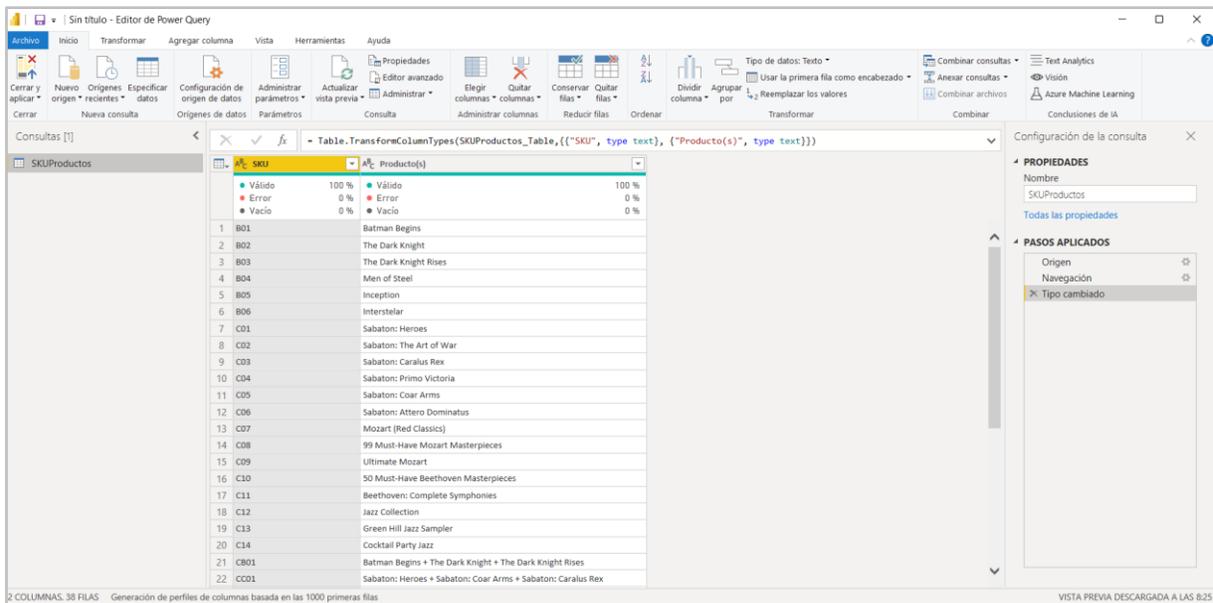


Figura 1. 8 – Ventana/Editor de Power Query con una Consulta

El editor de *Power Query* sigue ciertas reglas al igual que sus elementos que derivan del desarrollo en *Microsoft*, conocerlas nos brindará la libertad de movernos con facilidad y seguridad en *Power BI*, manipularlo como un todo y manejar cada elemento con gran solvencia.

Entrelazamiento

El **Entrelazamiento** del editor de *Power Query* señala que: para cada archivo de *Power BI* siempre hay asociado una ventana para trabajar con *Power Query*, por lo anterior:



¡El editor de Power Query siempre se llamará igual que su archivo de Power BI matriz!

Tanto el editor de *Power Query* como la interfaz analítica de *Power BI* en la barra de títulos siempre tienen el mismo nombre como prefijo, seguido del nombre *Power BI* o *Editor de Power Query*.

Lo anterior permite entrever que podemos tener varias ventanas o editores de *Power Query* abiertas.

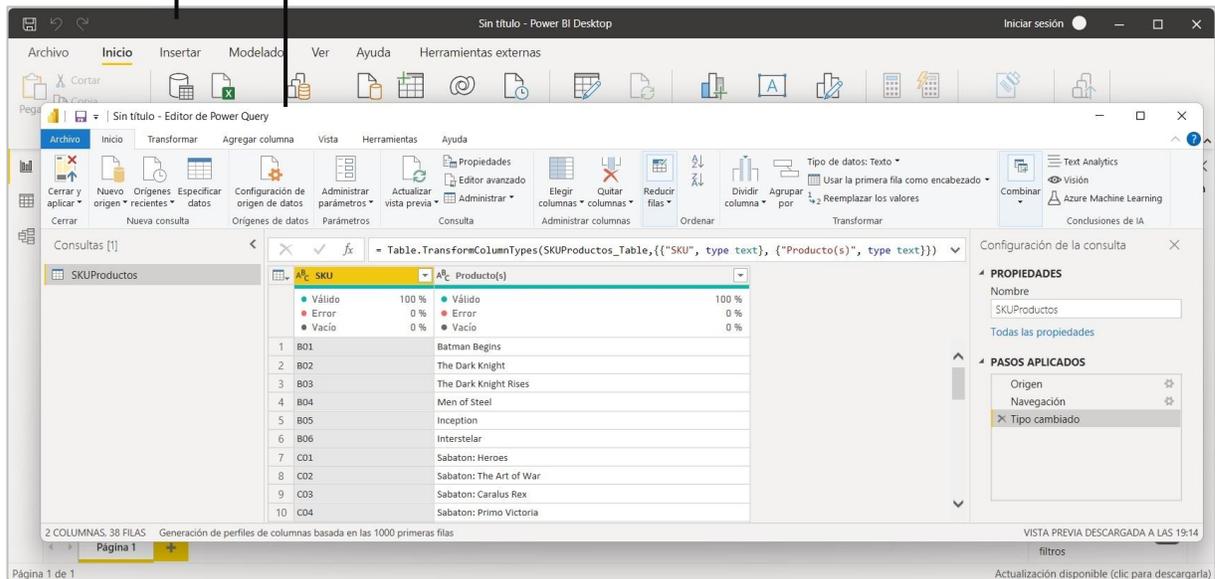


Figura 1.9 - Editor de Power Query Desktop e Interfaz Analítica de Power BI Desktop

La Interfaz de Power Query de Escritorio (Desktop)

Con la idea que: *Una consulta realmente está constituida por líneas de código del Lenguaje M* rondando por nuestra mente, al mismo tiempo puede surgir la pregunta *¿Power Query no es muy complicado de utilizar, considerando que debo conocer el Lenguaje M para emplearlo?*



— La respuesta querido(a) lector(a) es un: NO rotundo.

Todo ello se debe al *Editor de Consultas* o *Editor de Power Query*, lugar donde la magia ocurre, donde convertir datos sin forma alguna en una tabla bien estructurada con un par de clics no es una exageración.

¿Qué es el Editor de Power Query?

Es una Interfaz gráfica que nos permite agregar, editar o eliminar pasos a la consulta actual.

¡Toma Nota!



El editor de *Power Query* es como la grabadora de macros en Excel o las grabadoras de flujo web o de escritorio de *Power Automate Desktop*, que capturan nuestras acciones y los transforman en líneas de código o pasos.

Al final del día con el *editor de Power Query* aplicamos el proceso de *Extracción, Transformación y Carga de Datos*, lo que lo convierte en una potente herramienta **low code**, porque podemos aplicar y automatizar las tareas ETL sin tener experiencia en la creación de líneas de código.

El nombre de esta tecnología ha ido cambiando, hoy la conocemos como Power Query

NOTAS

Nota 1

Algunas de las opciones que se encuentran en *Power BI Desktop* pero que no están en el *Power Query* de *Excel* son las conclusiones IA, que permiten hacer *análisis de texto*, *visión* y *machine learning*.

Por su parte *Power Query* en *Excel* tiene opciones únicas, incluso contrastado con *Power Query Online*, como, por ejemplo: *Creación de tipo de datos*.

Nota 2

Son muy pocos comandos los que tienen ubicaciones ligeramente diferentes, por ejemplo: *Introducir datos* o que la creación de *Parámetro nuevo* en *Excel* no aparece con clic derecho en el *panel de campos*.

Interfaz de *Power Query Desktop: Power BI de Escritorio y Microsoft Excel*

El editor de *Power Query* es prácticamente el mismo en *Excel* y *Power BI desktop*: cambia el color, las opciones de carga, algunas transformaciones disponibles y la ubicación de ciertos comandos, pero en esencia es la misma interfaz con dos armaduras diferentes.

Una gran ventaja de *Power Query* en *Power BI* respecto al de *Excel*, deriva de las actualizaciones mensuales del primero, que también introduce mejoras en su parte de preparación de datos, aunque *Power Query Online* siempre es el primero en obtener todas las actualizaciones y mejoras.

Interfaz de *Power Query Online: Gen1: Power BI Service, Power Apps, Power Automate, Azure Data Factory, Azure Synapse Analytics y Dynamic 365 Customer Insights. Gen2: Microsoft Fabric con Synapse Data Factory*

El editor de *Power Query* online es una experiencia muy similar a la de escritorio, pero puede tener sus diferencias marcadas, dado que todas las actualizaciones se ven primero reflejadas aquí, tanto así que la experiencia es mucha más enriquecida en la nube, contando con funcionalidades y características únicas.

Hagamos un desglose de la interfaz gráfica, sin tener en cuenta aquellos elementos que se mantienen más comúnmente entre aplicaciones como: *Barra de Títulos*, *Ícono de cerrar*, *Ícono de maximizar*, *Ícono de minimizar*, *Botón de Ayuda* y *la Barra de Estado*.

Interfaz de Power Query

En Power BI Desktop

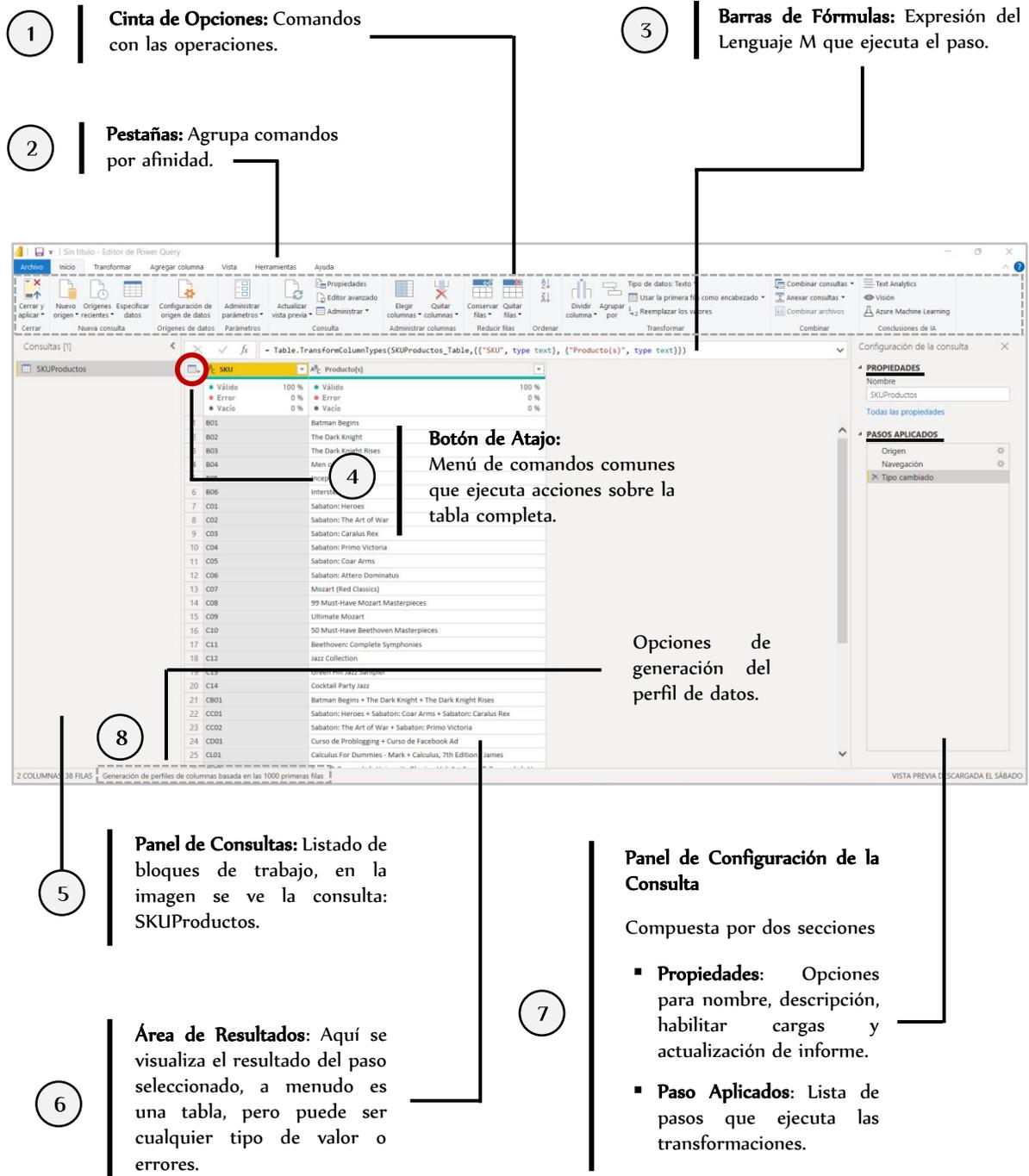


Figura 1. 10 – Editor de Power Query en Power BI de Escritorio (Desktop)

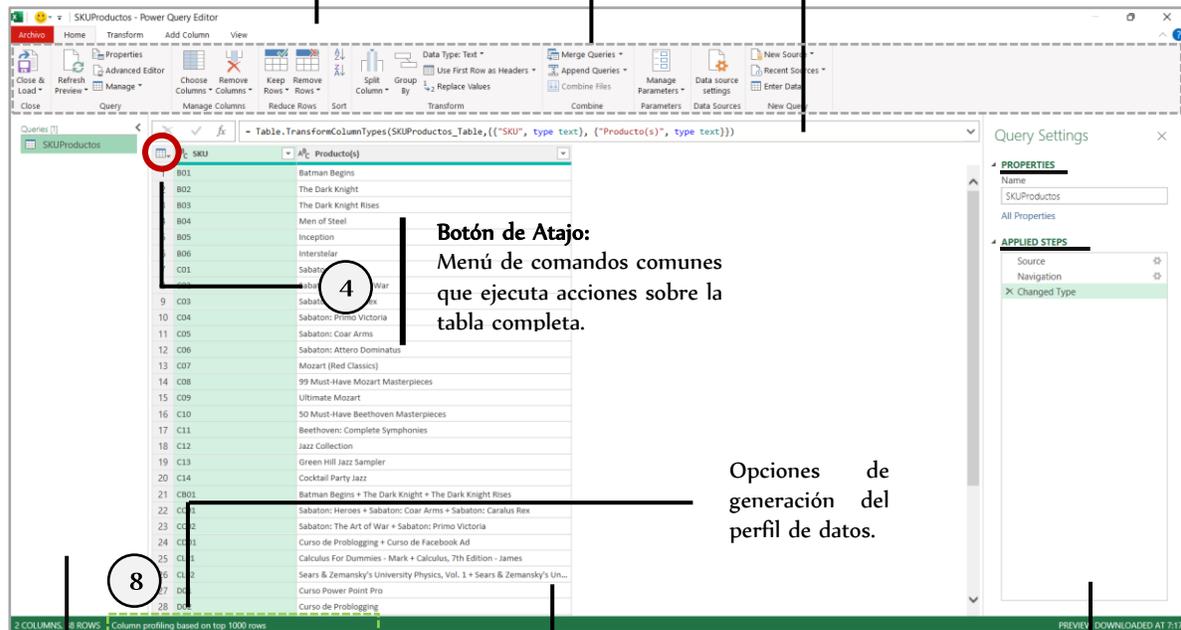
Interfaz de Power Query

En Microsoft Excel

1 **Cinta de Opciones:** Comandos con las operaciones.

3 **Barras de Fórmulas:** Expresión del Lenguaje M que ejecuta el paso.

2 **Pestañas:** Agrupa comandos por afinidad.



4 **Botón de Atajo:** Menú de comandos comunes que ejecuta acciones sobre la tabla completa.

Opciones de generación del perfil de datos.

5 **Panel de Consultas:** Listado de bloques de trabajo, en la imagen se ve la consulta: SKUProductos.

Panel de Configuración de la Consulta

Compuesta por dos secciones

- **Propiedades:** Opciones para nombre, descripción, habilitar cargas y actualización de informe.
- **Paso Aplicados:** Lista de pasos que ejecuta las transformaciones.

6 **Área de Resultados:** Aquí se visualiza el resultado del paso seleccionado, a menudo es una tabla, pero puede ser cualquier tipo de valor o errores.

7

Figura 1. 11 – Editor de Power Query en Microsoft Excel (Office 365)

1. **Cinta de Opciones:** La cinta de opciones es donde se encuentran todos los comandos de trabajo agrupados por afinidad, y estos a su vez están subdivididos por grupos (*etiquetas de la parte inferior*).

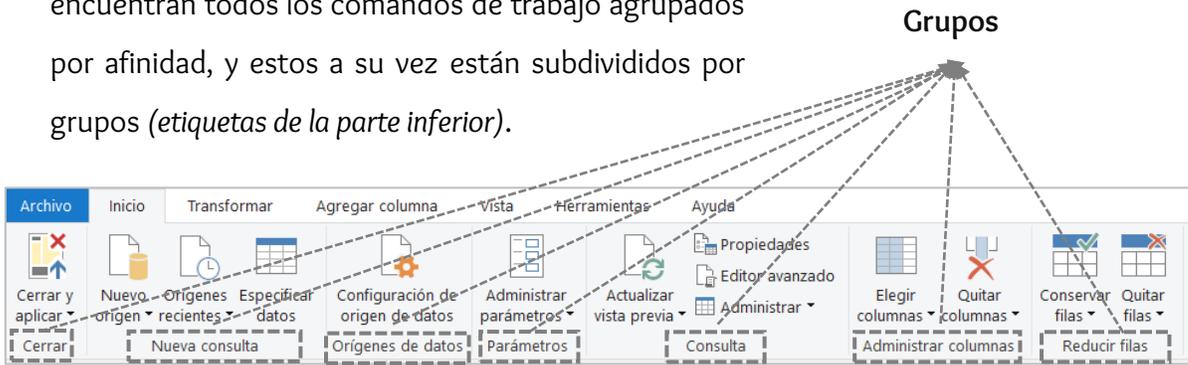


Figura 1. 12 – Cinta de Opciones en Pestaña Inicio y sus Grupos (Parte Inferior)

2. **Pestañas:** Estos grandes nombres permiten navegar entre los distintos grupos existentes en la cinta de opciones, en *Power Query* concretamente contamos con las siguientes pestañas:

- a. (Archivo)
- b. Inicio
- c. Transformar
- d. Agregar Columna
- e. Vista
- f. Herramientas
- g. Ayuda

NOTA

La pestaña Archivo es exclusiva para *Power Query Desktop*.

Por otra parte, el número de pestañas puede variar dependiendo de la acción que estemos ejecutando, a estas se les denomina pestañas contextuales, algunas de ellas:

- Consulta
- Transformar (Adicional)

3. **Barra de Fórmulas:** Muestra la expresión del *Lenguaje M* asociada al paso en el cual nos encontramos posicionados, esta barra de fórmulas por defecto está deshabilitada. [*Vista* → *Diseño* → *Barra de Fórmulas*]



4. **Botón de Atajo:** Es un menú con comandos comunes para acciones en la tabla.

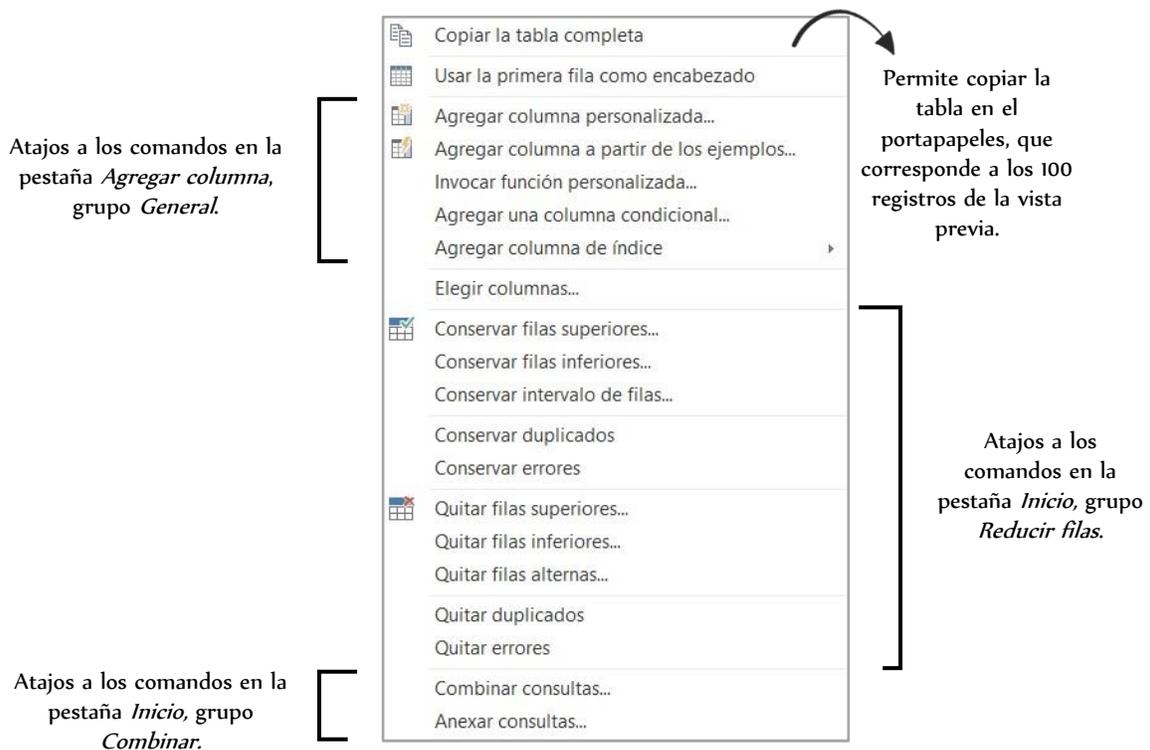
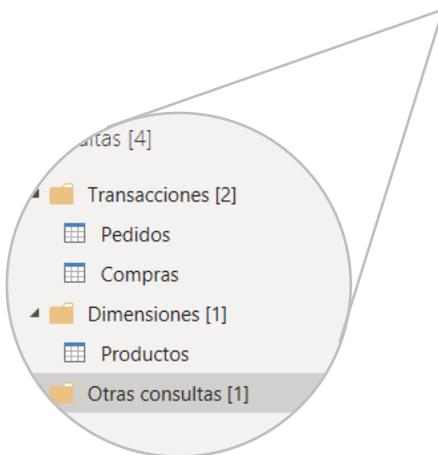


Figura 1.13 – Botón de Atajo, Menú con Comandos en Acciones Sobre la Tabla



Vistazo al Panel de Consultas

5. **Panel de Consultas:** Muestra las distintas consultas o bloques de trabajo que tiene el proyecto para navegar entre sí, como estos varían según su resultado y objetivo, podemos tener variedades de íconos. Además, podemos crear carpetas (*grupos*) para agrupar consultas. (*Clic derecho en el panel → Nuevo Grupo*).

6. **Área de resultados:** Como su nombre lo indica expone la salida (*resultado*) obtenido por la expresión, dependiendo de la consulta puede variar según el paso en el que estemos situados.

7. **Panel de Configuración de la Consulta:** Revela todas las propiedades inherentes a la consulta seleccionada, está subdividida en dos partes, áreas o secciones:

a. **Propiedades:** En esta parte vemos dos elementos a su vez:

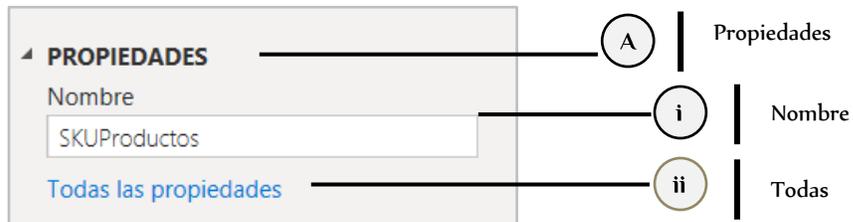


Figura 1. 14 – Sección Propiedades

i. **Nombre:** Aquí rebautizamos la consulta si así lo queremos.

ii. **Todas las Propiedades:** Aquí se muestran propiedades adicionales.

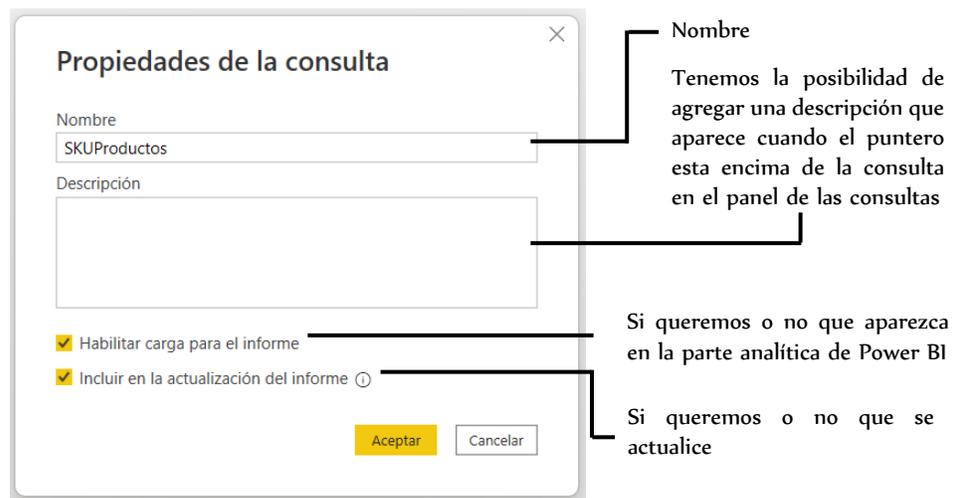


Figura 1. 15 – Propiedades de la Consulta

- b. **Pasos Aplicados:** Las fotos (*snapshots*) de cada operación que realizamos con *Power Query* se ven reflejados aquí como un paso en particular, de hecho, lo que visualizamos es el nombre o identificador de la expresión M, por consiguiente, si queremos ver un estado anterior en nuestro proceso, sólo es suficiente con pulsar clic en el nombre correspondiente.

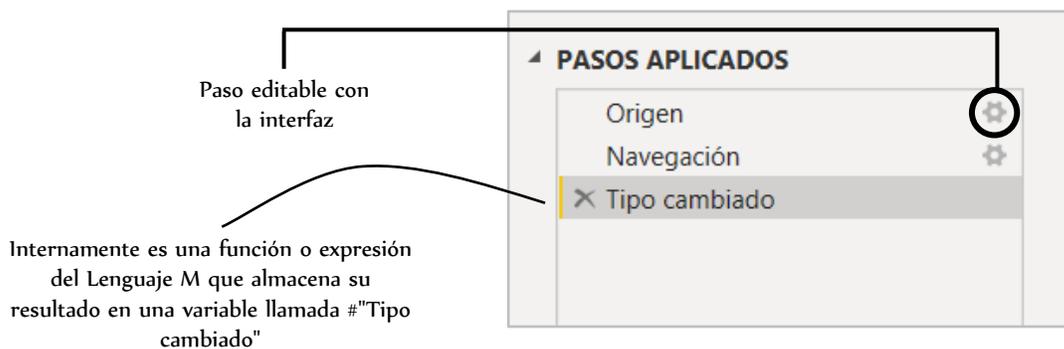


Figura 1. 16– Sección de Pasos Aplicados

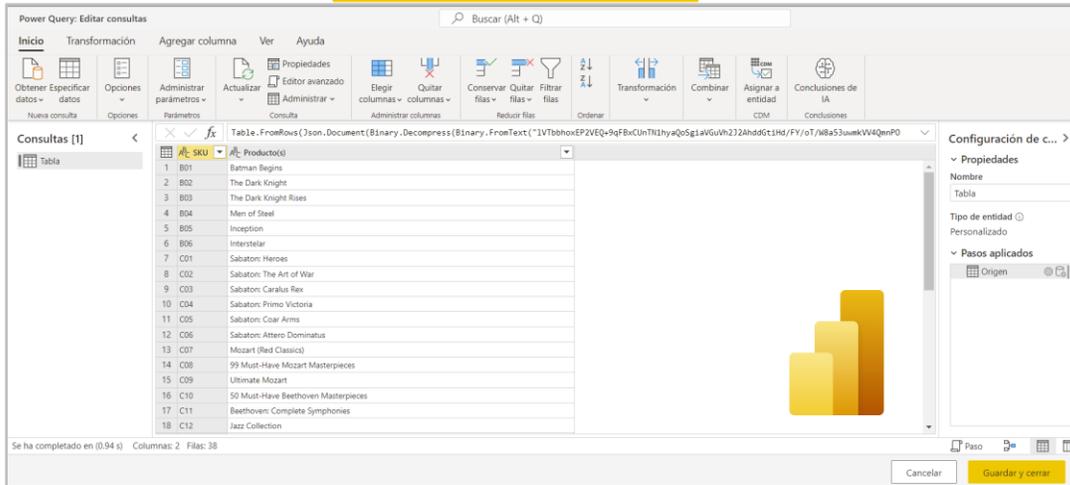
La Interfaz de Power Query Online: Dataflows Gen1

En *Power Query online* existe dos generaciones: *Flujos de datos Gen1* y *Flujos de datos Gen2*. Conversemos sobre la generación 1 primero. La interfaz de *Power Query online Gen1* es muy similar a la de *Power Query de escritorio*, por lo que una vez ubicado en el editor te vas a sentir en casa, recordemos que *Power Query online Gen1* se encuentra en:

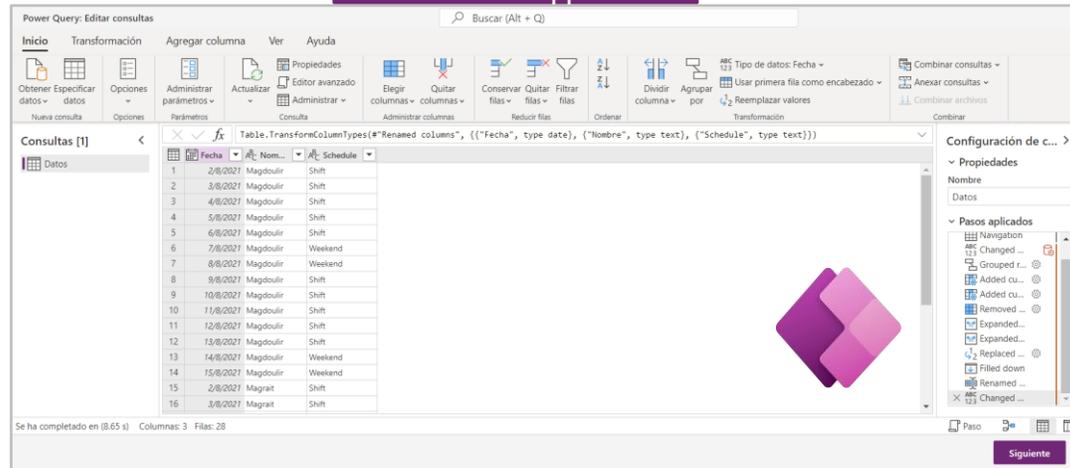
- Power BI Service (*Dataflows y Dataflows en datamarts*)
- Power Apps
- Power Automate (*Conector de SQL Server*)
- Azure Data Factory
- Azure Synapse Analytics
- Dynamic 365 Customer Insights

— Así luce *Power Query* en algunos de estos servicios:

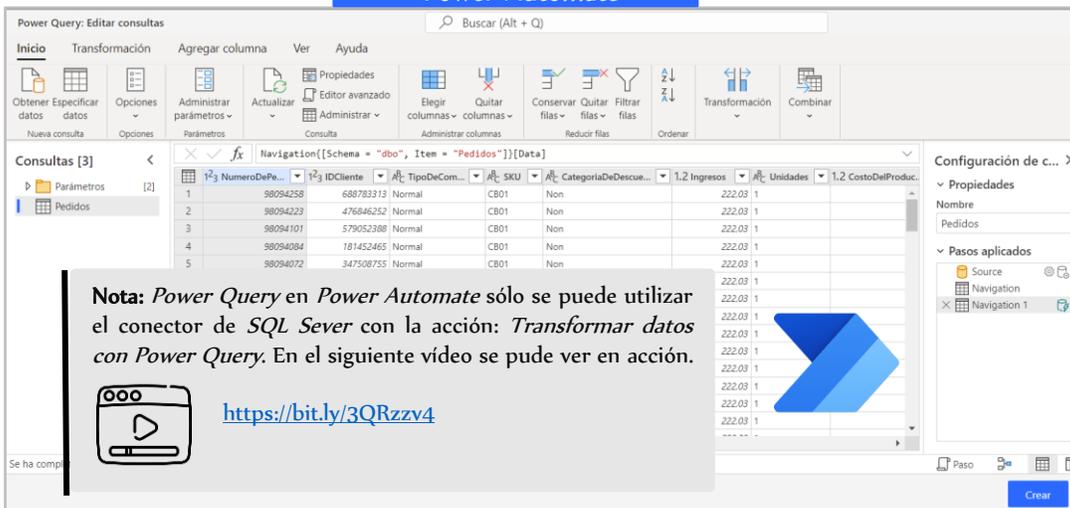
Power BI Service



Power Apps



Power Automate



Power Query online Gen1 tiene diferencias con *Power Query de escritorio*, las más sutiles están asociadas a la ubicación de algunos comandos y a la forma de "cerrar y aplicar", ya que en *Power BI Service*, *Power Apps* y *Power Automate* se resumen en el botón *Guardar y cerrar*, *Siguiente* o *Cerrar*, que se localizan en la parte inferior derecha, no obstante, también es cierto que *Power Query online* es superior a *Power Query de escritorio*, porque está dotado con funcionalidades y características únicas, a decir verdad, el propio equipo de producto ha dicho y pone sus esfuerzo en la versión en la nube.



NOTA

Algunos elementos adicionales y distintivos de *Dataflows Gen1 y Gen 2* son:

- El buscador de comando en la barra de título
- Las representaciones visuales de los tipos de combinación
- La opción de marcar como clave
- Contador de ejecución

Empero, todas estas características se compartirán con *PQ Desktop* desde febrero del 2024.

— Las siguientes funcionalidades se compartirán entre *PowerQuery de Escritorio* y *Power Query Online Gen1 y Gen 2* desde febrero 2024:

- Vista de esquema
- Vista de diagrama
- Script (de la consulta)
- Plan de consulta
- Control visual de plegado de consultas

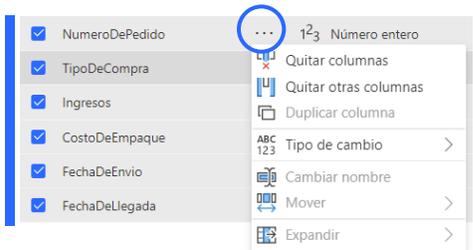
Vista de Esquema

En la pestaña *Ver*, grupo *Vista previa* notaremos el comando *Vista de esquema*, si pulsamos clic encima veremos:

| Nombre | Tipo | Clave |
|----------------|----------------|-------|
| NumeroDePedido | Número entero | |
| TipoDeCompra | Texto | |
| Ingresos | Número decimal | |
| CostoDeEmpaque | Número entero | |
| FechaDeEnvio | Fecha/Hora | |
| FechaDeLlegada | Fecha/Hora | |

Figura 1. 17 – Vista de Esquema en Power Query Online

Con la *vista de esquema* activada observamos como cambia el *Área de resultados* para listar las columnas de la tabla, su tipo de datos y si es una columna clave, sin embargo, lo más importante es que brinda la posibilidad de manipular una o más columnas de una manera más amigable, por ejemplo, al pulsar en los tres puntos (...) contaremos con opciones para: Quitar columna, Duplicar columna, Cambiar el tipo de dato, Cambiar el nombre, Mover y Expandir, lo estupendo es que estas acciones también se pueden ejecutar en lote, en otros términos, a varias columnas en simultaneo.



Vista de Diagrama

En la pestaña *Ver*, grupo *Diseño* notaremos las opciones de *Vista de diagrama*, si pulsamos clic encima veremos:

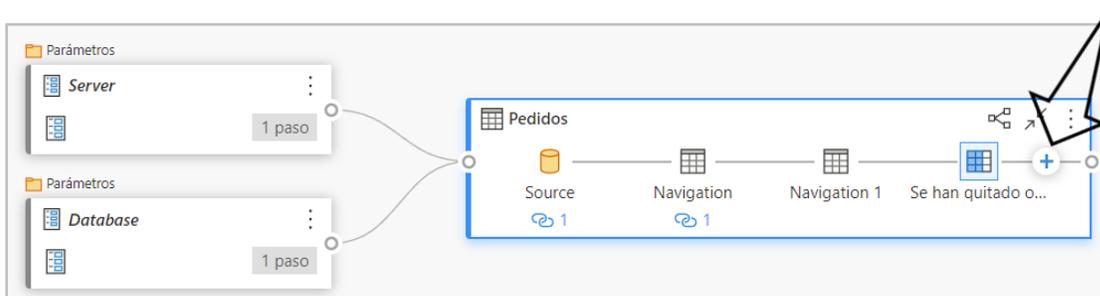


Figura 1. 18 – Vista de Diagrama en Power Query Online

La *vista de diagrama* es una experiencia totalmente diferente, la cual mediante cajas nos ofrece un mapa global de todo el flujo de datos, mostrando gráficamente las relaciones existentes entre consultas, así como los vínculos en los pasos de transformación, por si fuera poco, la vista de diagrama nos permite modificar y realizar las trasformaciones directas, por ejemplo: si una transformación posterior vamos a remover filas en blanco, entonces, pulsamos clic en el botón (+) en el extremo de la caja representativa de la consulta:



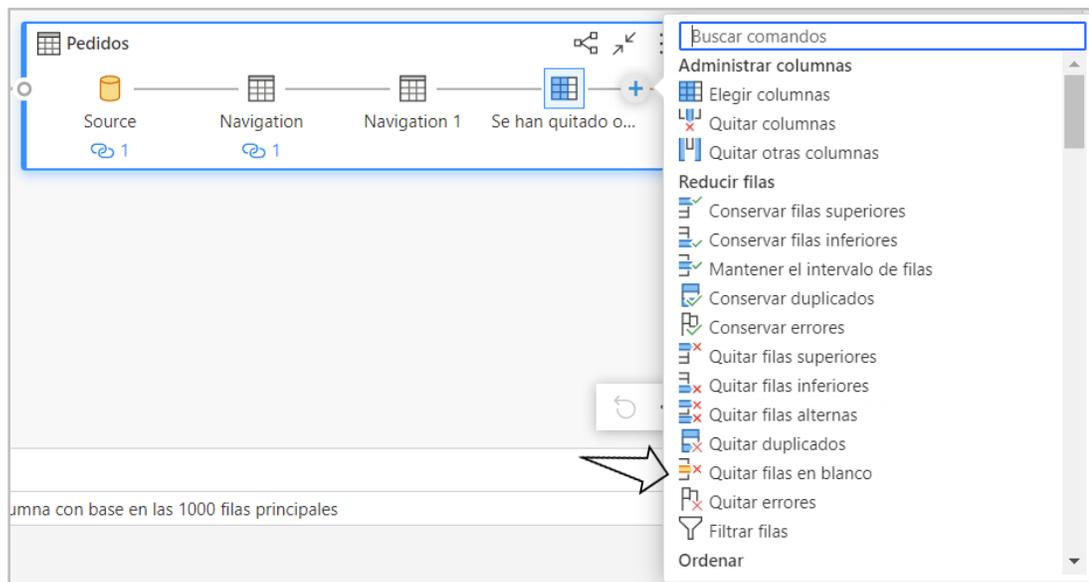


Figura 1. 19 – Menú de Transformaciones en la Vista de Diagrama

En este listado de comandos encontraremos la opción: *Quitar filas en Blanco*.

Script (de la Consulta)

Resulta que las transformaciones aplicadas desde la interfaz se transforman en código, conocido como expresiones del *Lenguaje M*. Por defecto, en *Power Query Online* y *Power Query de escritorio* la visualización de estas expresiones está desactiva.

Para activarlo en *Power Query Online* nos dirigimos a la pestaña *Ver*, grupo *Diseño*, desplegamos las opciones de *Script* y pulsamos clic en *Script de Pasos*.

The image shows the Power Query script editor with a table view. The table has the following data:

| NumeroDePedido | IDCiente | TipoDeCompra | SKU | CategoriaDeDescuento | Ingresos | Unidades | CostoDelProducto |
|----------------|-----------|--------------|------|----------------------|----------|----------|------------------|
| 98094258 | 688783313 | Normal | CB01 | Non | 222.03 | 1 | 7.99 |

Figura 1. 20 – Script de Paso

Sin embargo, esto nos mostrará sólo la expresión del paso actual seleccionado.

En *Power Query Online*, tenemos la posibilidad de ver las expresiones de todos los pasos, esto se logra de igual forma, es decir, pestaña *Vista*, Grupo *Diseño*, Opciones de *Script*, empero, aquí optamos por la posibilidad de *Script de la Consulta*.

The screenshot shows the Power Query Online interface. At the top, there is a DAX query script:

```

1 let Source = Sql.Databases(Server), Navigation = Source[Name = Database][Data],
2 #"Navigation 1" = Navigation[Schema = "dbo", Item = "Pedidos"][Data],
3 #"Se han quitado otras columnas." = Table.SelectColumns(#"Navigation 1", {"NumeroDePedido", "IDCliente", "TipoDeCompra", "SKU",
   "CategoriaDeDescuento", "Ingresos", "Unidades", "CostoDelProducto", "CostoDeEnvio", "CostoDeEmpaque", "FechaDePedido", "FechaDeEnvio",

```

Below the script, there is a table with columns: NumeroDePedido, IDCliente, TipoDeCompra, SKU, CategoriaDeDescuento, Ingresos, Unidades, CostoDelProducto. The first row of data is:

| NumeroDePedido | IDCliente | TipoDeCompra | SKU | CategoriaDeDescuento | Ingresos | Unidades | CostoDelProducto |
|----------------|-----------|--------------|------|----------------------|----------|----------|------------------|
| 98094258 | 688783313 | Normal | CB01 | Non | 222.03 | 1 | 7.99 |

Figura 1. 21 – Script de Consulta

Plan de Consulta

Es otra característica exclusiva de *Power Query Online* sólo hasta febrero del 2024, la cual permite tener una vista más profunda de lo que ocurre internamente para evaluar la consulta, y es bastante útil para detectar porque el plegado de consultas se daña.

¡Toma Nota!



Power Query hace todo lo posible por delegar el trabajo de transformación al origen, por ejemplo: al remover una columna en *Power Query*, este no carga la tabla en memoria para ejecutar la transformación de forma local, sino que toma la operación expresada en *Lenguaje M* que ha demandado el usuario y la traduce al idioma nativo del origen: *SQL*, para que así sea este el que haga el trabajo pesado, aprovechando las bondades de procesamiento y optimización del origen, con ello *Power Query* recibe el resultado ya transformado ganando tiempo, a este se le denomina: *Plegado de Consultas*.

No todos los orígenes se pueden *Plegar*, por ejemplo, con *Excel* no es viable, ya que no tiene un lenguaje de consultas, además, si el origen admite plegado, no quiere decir que todas las operaciones lo puedan hacer.

Para ver el plan de consultas basta con pulsar clic derecho encima del paso en el *panel de pasos aplicados* y el menú que se despliega seleccionar *Plan de consultas*.

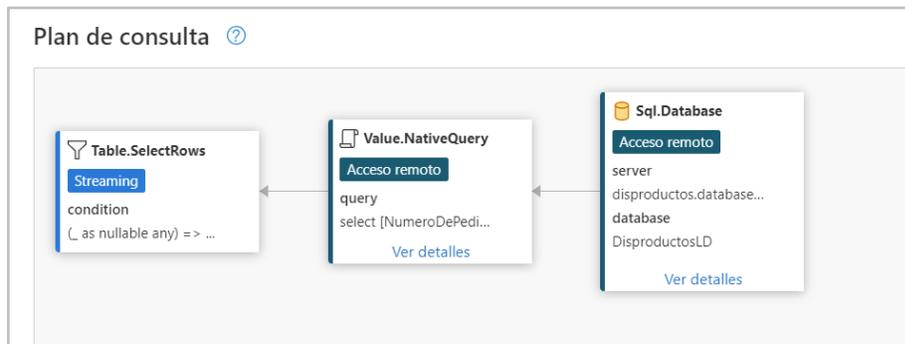


Figura 1. 22 – Plan de Consulta

Control Visual de Plegado de Consulta

Mediante un icono en el extremo derecho de cada paso, se señala que sucede con el plegado de la consulta.

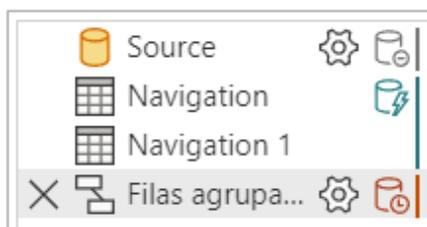


Figura 1. 23 – Indicador Visual

NOTA

Para conocer los detalles de los iconos de indicadores y su significado, léase el artículo:

<https://docs.microsoft.com/en-us/power-query/step-folding-indicators>

¿Cómo ir al Editor de Power Query Online Gen1?

Aunque hemos comentado algunas características distintivas de *Power Query Online (Gen 1)*, no hemos señalado cómo se accede a este editor en la nube, el punto aquí es que: el acceso a *Power Query Online (Gen 1)* varía un poco de acuerdo con el servicio que estamos utilizando, *Power BI*, *Power Apps*, etc.



Power Query Online Gen1 en Power BI Service

Ubicados en el servicio de *Power BI* en la nube, es necesario crear un área de trabajo “personalizada” (*Workspaces*), dado que los *flujos de datos* no pueden ser creados en el área de trabajo personal (*My Workspace*).

¡Toma Nota!



Al tener cómo única opción la utilización de *flujos de datos* en áreas de trabajo, quiere decir que *se necesita licencia de Power BI pro como mínimo para poder acceder, además, hay funcionalidades que requieren de licencia Premium o Premium por usuario.*

Una vez creado el área de trabajo, basta con ir a la parte superior, pulsar en *Nuevo* y seleccionar *flujo de datos*.

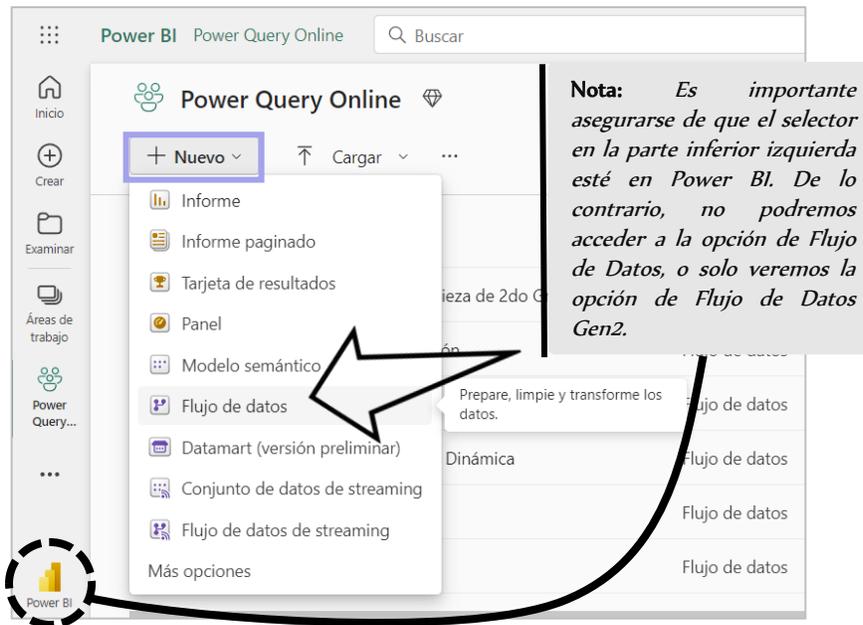


Figura 1. 24 – Crear Flujo de Datos en el Servicio de Power BI

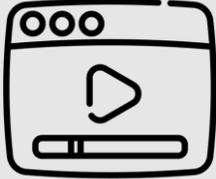
En la siguiente página es viable iniciar con *Definir tablas nuevas*.



Figura 1. 25 – Definir tablas nuevas en Flujo de Datos

NOTA

El siguiente vídeo muestra la instalación de una puerta de enlace a partir del minuto 4:36.



<https://bit.ly/3OFINck>

A partir de allí nos encontraremos con una experiencia “similar” a la versión de escritorio, puesto que expone una serie de *conectores* para extraer datos desde: *Excel*, *Texto o CSV*, *PDF*, *Carpeta*, etc. No obstante, una diferencia que se debe tener en consideración es que: si necesita extraer datos de archivos locales (*on-premise*) se debe procurar por contar una puerta de enlace (*Gateway*).

— *En este libro y sus diversos capítulos estudiaremos los pormenores los flujos de datos Gen1 y Gen2.*

¡Referencia Cruzada!



Otro libro que examina detalladamente los *flujos de datos de Gen1* como herramienta de autoservicio para ETL (*Power Query Online*) es *Power BI Dataflows*, del MVP Francisco Mullor.

Power BI Dataflows

ISBN: 9781737497844 (<https://bit.ly/3bmYtIT>)

Organización Jerárquica A/O

Power Query es de una generosidad desbordante que en cuestión de segundos y sin conocerlo mucho, nos permite realizar tareas impresionantes, no obstante, si tenemos unos conceptos adicionales seremos unos campeones.

Conceptos de «Fuego» | Acciones y Operaciones



*La organización jerárquica de acciones y operaciones ha sido nuestro esfuerzo en categorizar cada actividad que podemos realizar en Power Query, esto quiere decir que **NO es una terminología oficial de Microsoft**, no obstante, aún seguimos trabajando en ella para ofrecer una referencia y organización de las acciones y operaciones que brinda Power Query a nivel de interfaz, a continuación, veremos de que se trata.*



Acciones y Operaciones

En la segunda etapa del proceso ETL contamos con distintas operaciones que están enmarcadas en tres grandes núcleos denominados: «Acciones», cada una tiene un fin y es esencial conocer el objetivo de cada una de ellas:

Acción

- **LIMPIEZA DE DATOS:** Permite reajustar la forma de una tabla modificando su estructura a nivel de filas, columnas o incluso alterar toda la composición de la tabla para lograr el formato tabular óptimo, esta acción está compuesta por las siguientes operaciones:

| | |
|--|--|
| <ul style="list-style-type: none"> ○ Reducción <ul style="list-style-type: none"> ▪ <i>Exclusión de filas</i> ▪ <i>Quitar columnas</i> ▪ <i>Promover encabezado</i> ○ Transformación de Datos <ul style="list-style-type: none"> ▪ <i>Transformación a nivel de columnas</i> ▪ <i>Reemplazar valores</i> ▪ <i>Dividir columnas</i> ▪ <i>Asignación de nombres</i> ○ Transposición (podría estar abajo) ○ Rellenar | <p>Limpieza de Datos de Primer Grado:</p> <p><i>Técnicas de Depuración</i></p> |
| <ul style="list-style-type: none"> ○ Anulación de Dinamización (Unpivot) ○ Columna Dinámica (Pivot) ○ Agrupación | <p>Limpieza de Datos de Segundo Grado:</p> <p><i>Técnicas de Restructuración por Alteración</i></p> |

Operaciones y Sub-operaciones

De la primera acción: Limpieza de datos hay que destacar que consta de seis operaciones: *reducción*, *transformación*, *transposición*, *rellenar*, *anulación de dinamización y columna dinámica*, no obstante, con las dos primeras podemos bajar un poco más en la jerarquía y ver que están compuestas por *sub-operaciones*, dado que a este nivel ya corresponde a comandos u opciones concretas en la interfaz de *Power Query*.

A nosotros nos encanta hacer estas diferencias, puesto que brinda una estructura mental clara y precisa de cada aspecto tanto en el editor como en la práctica de preparación de datos, a pesar de ello, por simplicidad a lograr en el presente libro hacemos la siguiente salvedad.



¡Toma Nota!

Ten en cuenta que a las sub-operaciones las llamaremos tan sólo operaciones en algunas ocasiones a lo largo del texto, puede parecer una diferencia sutil, pero es lo que nos diferencia en preparación de datos inteligente.

Técnicas (Limpieza de Datos)

Otro punto es que las primeras cuatro grandes operaciones se han categorizado bajo el nombre de: **Limpieza de datos de primer grado: técnicas de depuración**, mientras que las tres últimas grandes operaciones bajo la categoría de: **Limpieza de datos de segundo grado: técnicas de reestructuración por alteración**.

Lo anterior es así para tener una distinción adicional de aquellas operaciones que cambian columnas o celdas concretas, de las que cambien la estructura completa de la tabla desde el principio.

Acción

- **CONSOLIDACIÓN DE DATOS:** Consiste en fusionar dos o más tablas para conseguir una gran tabla maestra con todos los datos que sean de interés para el análisis, la forma de pegar las tablas depende del análisis que deseemos lograr y de la composición de las tablas, esta acción está compuesta por las operaciones:

- **Anexar (Append)**
 - **Combinar (Merge)**
 - *Externa izquierda*
 - *Externa derecha*
 - *Externa completa*
 - *Asignación de nombres*
 - *Interna*
 - *Anti izquierda*
 - *Anti derecha*

Acción

- **ENRIQUECIMIENTO DE DATOS:** Consiste en dotar una tabla con registros creados artificialmente con métodos estadísticos y de minería de datos para enriquecer la muestra de datos.

- **Enriquecimiento Horizontal**
 - **Enriquecimiento Vertical**

¡Toma Nota!



Construcciones sencillas de columnas y filas pueden tomarse como enriquecimiento horizontal, no obstante, la verdadera “magia” de esta acción están en enriquecer la muestra de datos con estadísticas, esto requiere a su vez del *Lenguaje M*.

Extracción (Conectores)

Si no hay datos no es viable hacer mayor cosa, por esta razón debemos conocer como conectarnos o extraer datos de diversos orígenes para que alimenten *Power Query*.

La gama de orígenes que pueblan *Power Query* dista de ser insignificante, de hecho, existen tantas tonalidades que podemos subdividir en 5 grupos:

- **Archivos:** Orígenes de Datos Estándar
- **Bases de Datos:** Orígenes de Bases de Datos Relacional
- **Microsoft Fabric:** Orígenes de la Plataforma en era IA
- **Servicios en Línea:** Orígenes en la Nube
- **Power Platform:** Orígenes del Ecosistema Platform
- **Azure:** Orígenes de Diversidad de Servicios de Azure
- **Otras:** Otros orígenes sin Categoría Particular

A un origen singular, por ejemplo: *Excel* se le denomina conector, y si bien el catálogo crece constantemente, puede que necesitemos uno que no se encuentre presente, por ello *Power BI* en general es tan potente que permite crear nuestros propios conectores personalizados para *Power Query*.



¡Referencia Cruzada!



La creación de conectores personalizada es una temática avanzada que requiere de mucha programación y herramientas externas como *Visual Studio*, aunque se pueden hacer algunos sencillos, su profundización es exigente, es por este motivo que sólo es tratado en el libro #3 de la trilogía de *Power Query* y el *Lenguaje M*.

Libro #3 | Guía Definitiva al Lenguaje M
ISBN

NOTA

En la categoría “otros” se encuentran como su nombre lo indica orígenes diversos que no encajan en ninguna de las clasificaciones predeterminadas, por ejemplo: ODBC, OLE DB, Google Sheet, Consulta en Blanco, etc.

Archivos: Orígenes de Datos Estándar

Conectarse a un origen de datos es sencillo, ya que todos siguen la posterior ruta en escritorio:

Inicio → Datos Externos → Obtener Datos → ...

Nos referimos a la “ruta” anterior para extraer datos de un origen de datos como la “ruta dorada” desde la interfaz de *Power BI*, sin embargo, si nos encontramos en la interfaz de *Power Query* la ruta es:

Inicio → Nueva Consulta → Nuevo Origen → ...

En esta categoría enmarcamos esos conectores más utilizados y comunes.

Tabla 1.1 Archivos: Orígenes de Datos Estándar

| Conector | Extensión |
|-------------------|-------------|
| Libro de Excel | .xlsx |
| Texto o CSV | .txt - .csv |
| Carpeta | - |
| Xml, JSON | .xml, .json |
| PDF | .pdf |
| Parquet | .parq |
| Especificar Datos | - |

Conector Excel: Guía Definitiva

Para conocer los detalles de lo que es el conector más popular (*Excel*) utilizaremos la tabla en el siguiente archivo:

¡Let's Play!



ARCHIVO DE TRABAJO: *En el compendio de archivos y carpeta Capítulo 3, puedes encontrar el archivo de Excel con nombre: CAPIEJ2-Ventas_Pedidos.xlsx. (Lo abrimos para un vistazo)*



Ruta Dorada
Power BI Desktop



Ruta Dorada Power
Query Desktop

| Solicitado Por: Ing. Miguel Caballero | | | | Fecha de Solicitud: 11 de noviembre de 2015 | | | | Tipo de Confidencialidad: Global | | | |
|---------------------------------------|----------|------------------------|--------------------|---|--------------------|----------------|----------|----------------------------------|--|--|--|
| Número de Pedido - Tip. Compra | SKU | Categoría de Descuento | Ingresos | Unidades | Costo del Producto | Costo de Envío | | | | | |
| Total Resumen: | | | | | | | | 133.41712012373816 | | | |
| **** | **** | **** | **** | **** | **** | **** | **** | | | | |
| **** | **** | **** | **** | **** | **** | **** | **** | | | | |
| **** | **** | **** | **** | **** | **** | **** | **** | | | | |
| #iDIV/0! | #iDIV/0! | #iDIV/0! | #iDIV/0! | #iDIV/0! | #iDIV/0! | #iDIV/0! | #iDIV/0! | | | | |
| xxxx | xxxx | xxxx | xxxx | xxxx | xxxx | xxxx | xxxx | | | | |
| 98094261Normal | CC01-2 | Non | 105.06 | 2 | 8,55 | 4,55 | 4,55 | | | | |
| 98094260Normal | L01-5 | Non | 61.5 | 2 | 8,25 | 3,25 | 3,25 | | | | |
| 98094258Normal | CB01-2 | Non | 222.02999999999997 | 1 | 7,99 | 4,99 | 4,99 | | | | |
| 98094259Devolución | CB01-2 | Non | 666.08999999999992 | 3 | 5,61 | 5,61 | 5,61 | | | | |
| 98094257Normal | CC02-1 | Non | 80.16 | 3 | 9,99 | 3,99 | 3,99 | | | | |
| 98094256Normal | L01-5 | Non | 30.75 | 1 | 5,14 | 6,14 | 6,14 | | | | |
| 98094255Normal | CB01-3 | Non | 666.08999999999992 | 3 | 7,14 | 4,14 | 4,14 | | | | |
| 98094253Normal | L01-4 | Non | 61.5 | 2 | 8,58 | 3,58 | 3,58 | | | | |
| 98094254Normal | L02-1 | Non | 120.60000000000001 | 3 | 9,55 | 5,55 | 5,55 | | | | |
| 98094250Normal | CC01-4 | Non | 105.06 | 2 | 7,68 | 3,68 | 3,68 | | | | |
| 98094251Normal | CC01-7 | Non | 52.53 | 1 | 8,99 | 5,99 | 5,99 | | | | |
| 98094252Devolución | CC02-7 | Non | 53.44 | 2 | 7 | 5 | 5 | | | | |
| 98094244Normal | B06-1 | Non | 50.36 | 1 | 7 | 5 | 5 | | | | |
| 98094247Devolución | CC01-3 | Non | 52.53 | 1 | 8,1 | 6,1 | 6,1 | | | | |
| 98094248Normal | CC02-3 | Non | 53.44 | 2 | 7 | 3 | 3 | | | | |
| 98094249Devolución | CC02-6 | Non | 80.16 | 3 | 5,58 | 5,58 | 5,58 | | | | |

Figura 1. 26- Fragmento de la Tabla Pedidos (La tabla esta acortada en el extremo derecho y en la parte inferior)

La tabla *Pedidos* es interesante para tratar todos los temas de limpieza de datos de primer grado, sin embargo, esto será tema de un capítulo posterior, por lo pronto sigamos los pasos mencionados a continuación:

1. Abrimos un archivo nuevo de *Power BI*.
2. Realizamos la ruta dorada, explícitamente: Vamos a la pestaña *Inicio*, grupo *Datos Externos* y desplegamos las opciones comando *Obtener datos*.

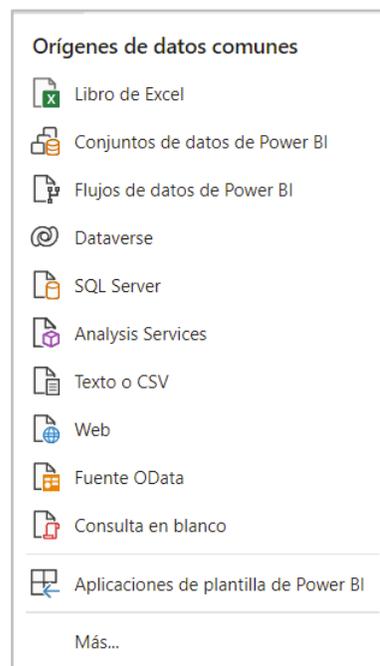


Figura 1. 27 – Opciones de Obtener Datos

3. Pulsamos clic encima de la opción *Libro de Excel*.
4. Cuando navegamos por nuestro computador solo nos mostrará carpetas y archivos de Excel, ubicamos el de *CAP1EJ2-Uentas_Pedidos.xlsx*

¡Riesgo: Advertencia!



El conector *Excel* permitirá visualizar las extensiones: *.xlsx*, *.xls*, *.xlsm* y *.xlsb*; sin embargo, los libros heredados (*legacy version*): *.xls* y *.xlsb* requieren del proveedor de access OLEDB (o ACE) para poder ser leídos, de lo contrario devuelve un error: *'Microsoft.ACE.OLEDB.12.0'*. Para descargar el proveedor léase el artículo: <https://bit.ly/3NOZqkp>

Los archivos *.xlsx* y *.xlsm* los carga correctamente sin ninguna instalación previa. Cualquier otro tipo de extensión asociada a *Excel* como: *.xls*, *.xlsxt*, *.xlst*, etc. Ni siquiera permitirá su visualización, entre otras cosas por compatibilidad y porque algunas son plantillas.

Lo siguiente que nos va aparecer es un cuadro de diálogo con el nombre de *Navegar*.

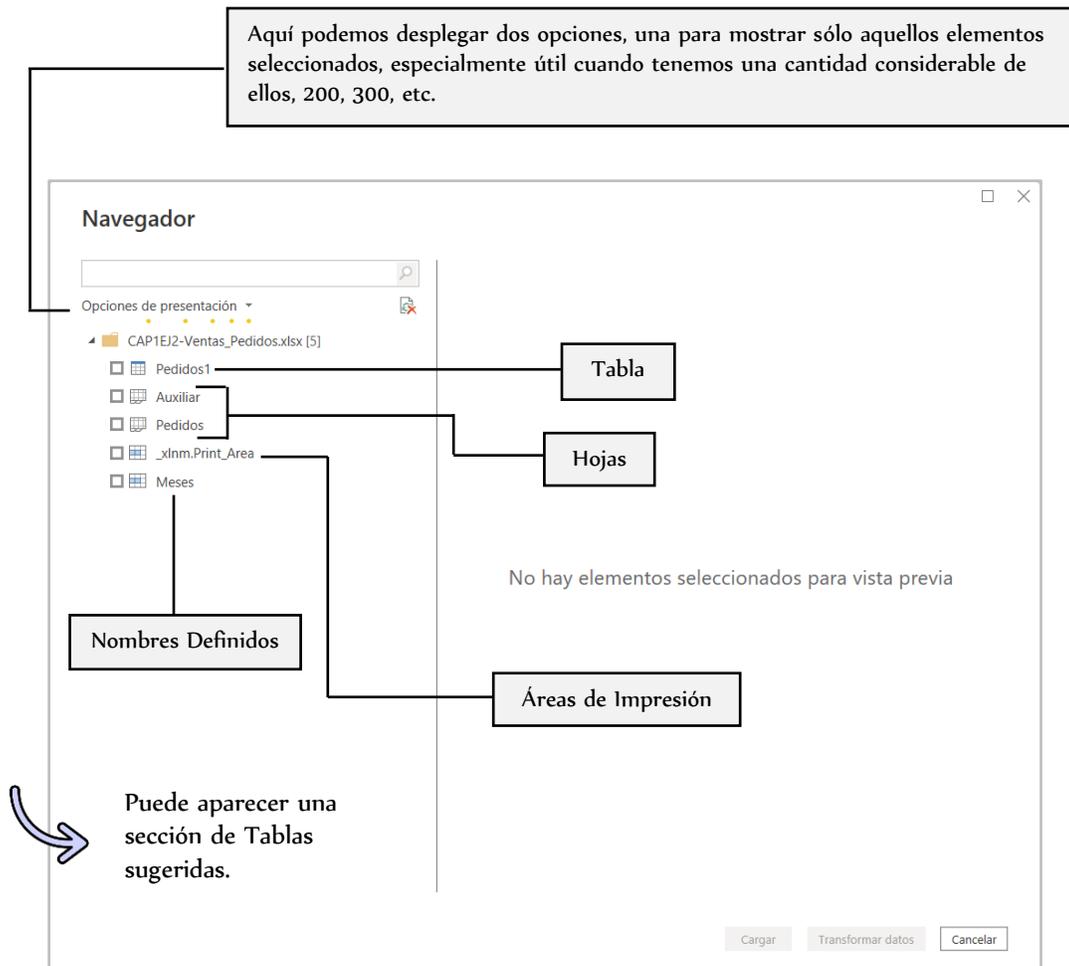


Figura 1. 28 – Cuadro de Diálogo Navegar

Como se puede observar en el cuadro de diálogo *Navegador* aparecen diversos elementos:

- Pedidos1
- Auxiliar
- Pedidos
- _xlnm.Print_Area
- Meses

– ¿Qué es cada elemento?

Para dilucidar la lista de elementos es imperativo conocer que cuando cargamos un archivo de *Excel* este nos mostrará lo siguiente:

- Tablas Estructuradas
- Hojas
- Nombres Definidos
- Áreas de Impresión

Para distinguirlos en la interfaz de *Power BI* vemos en el extremo izquierdo de cada elemento un ícono representativo.

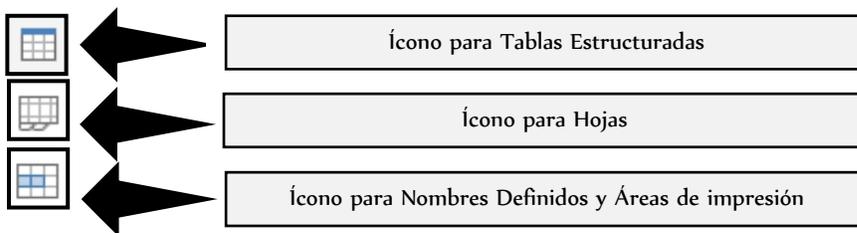


Figura 1. 29 – Íconos de Conector de Excel: Tablas, Hojas y Nombres Definidos

Para lo que necesitamos es claro: los datos de la tabla *Pedidos*, no obstante, en dicho sentido tenemos dos posibilidades: *Pedidos1* y *Pedidos*.

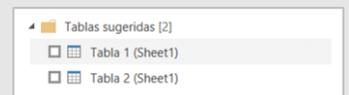
NOTA: Tablas Sugeridas

Se debe remarcar qué, si se conecta a un libro de *Excel* que no contiene específicamente una sola tabla, *Power Query* intentará crear una lista sugerida de tablas entre las que puede elegir.

Considere una hoja en un libro de *Excel* que contiene datos en los rangos: A1 a B3, D5 a F8 y de B10 a B13, como se ilustra a continuación:

| | A | B | C | D | E | F |
|----|----|-----------|---|------------|----------|--------|
| 1 | ID | Producto | | | | |
| 2 | | 1 Manzana | | | | |
| 3 | | 2 Pera | | | | |
| 4 | | | | | | |
| 5 | | | | Fecha | Unidades | Precio |
| 6 | | | | 01/01/2022 | 2 | 3 |
| 7 | | | | 02/01/2022 | 3 | 4 |
| 8 | | | | 03/01/2022 | 1 | 4 |
| 9 | | | | | | |
| 10 | | Meses | | | | |
| 11 | | Enero | | | | |
| 12 | | Febrero | | | | |
| 13 | | Marzo | | | | |

Al conectarnos notaremos en la parte inferior izquierda del cuadro de diálogo navegador la carpeta: *Tablas sugeridas*, que entre corchetes encierra el número de tablas que sugiere, para este caso 2:

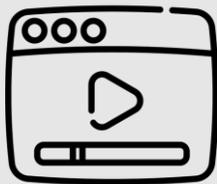


No siempre detectará todos los rangos, como lo es para el caso del rango B10 a B13, porque dependerá de los pasos de transformación que debe generar automáticamente mediante inteligencia artificial.

NOTA

Para *Power Query* de *Excel* 2013 y anteriores, también cargaba como opción los *Rangos de Filtros*, aquella funcionalidad de antaño que esgrimíamos con el comando *Avanzadas* en la pestaña *Datos* de *Excel*.

Si tienes curiosidad acerca de esta funcionalidad de *Excel*, el siguiente vídeo brinda una descripción breve.



<https://goo.gl/DxoqcB>

La opción de *Pedidos1* es la primera, como ya debieramos intuir corresponde a una tabla estructurada de *Excel* (aquellas que crean con *Ctrl + T*), mientras que la segunda hace referencia al conjunto de datos en la hoja como un todo.

— ¿Cuál de las dos opciones es mejor?

En este caso, la tabla estructurada, es decir, el primer elemento que en su ícono en el extremo izquierdo tiene encabezado azul será lo mejor, debido a que, si aparte de la tabla estructurada en la hoja en alguna otra parte existen más datos, entonces estos también serían cargados como partes de la consulta, alargando y complicando el proceso de limpieza.

— Por lo tanto:

5. Seleccionamos el elemento *Pedidos1*, si nuestra tabla no necesita limpieza podemos pulsar clic en el botón *Cargar*, pero si por el contrario hace falta transformación, pulsamos clic en el botón *Editar* para ir a la interfaz de *Power Query*.

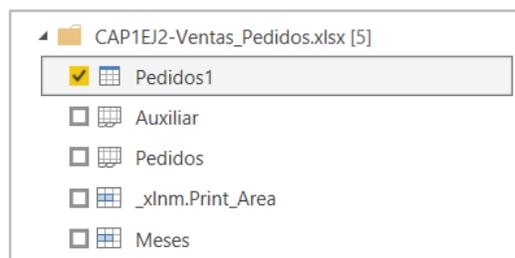


Figura 1. 30 – Seleccionar Consulta

Si optamos por *Editar*, varias cosas se podrán notar en la consulta recién creada, que es perfectamente válido destacar para una buena comprensión del punto de partida que tomaremos para la limpieza de datos.

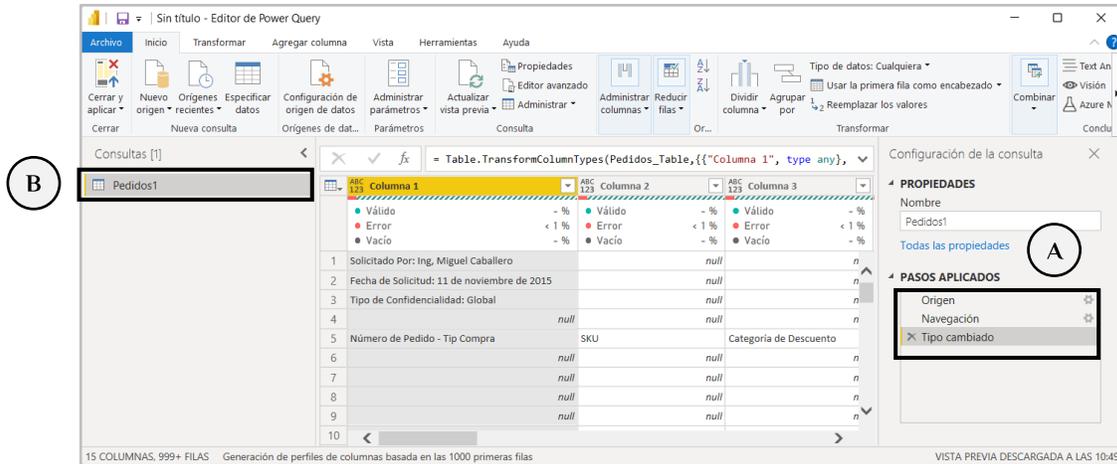


Figura 1. 31 – Tabla Pedidos Visualizada en el Editor de Power Query

En el *Panel Configuración de la Consulta* en su sección *Pasos Aplicados* se ve que de manera automática agrega tres pasos por defecto:

- Origen
- Navegación
- Tipo cambiado

Estos siempre serán un común denominador, y en el 80% de los casos aparecerán, en ciertas circunstancias puede incluir uno u otro paso adicional, pero los listados son lo que siempre persisten.

Por otra parte, en el panel *Consultas* se puede ver el bloque de trabajo.

NOTA

El paso automático de tipo cambiado viene habilitado por defecto, empero, este se puede deshabilitar en la pestaña *Archivo*, apartado *Opciones y configuración*, sección *Opciones* y en el cuadro de diálogo que se despliega en la sección de *Detección de tipos* se cambia a: *No detectar nunca los encabezados y tipos de columnas para los orígenes no estructurados*.

Excel se puede deshabilitar desde *Obtener datos: Opciones de consulta*.

¡Toma Nota!



En la imagen anterior se puede ver la barra de fórmulas que enseña las líneas del *Lenguaje M*, esta no se encuentra habilitada por defecto, para hacerlo debemos ir a la pestaña *Vista* y en el grupo *Diseño* habilitar la opción de *Barra de fórmulas*.

Podemos guardar el archivo de *Power BI* nuevo que hemos creado para reutilizarlo más adelante y agregar otras consultas

El primer paso para alcanzar este objetivo consiste en ir a la pestaña *Inicio*, grupo *Cerrar* y desplegamos las opciones de *Cerrar* y *Aplicar*.

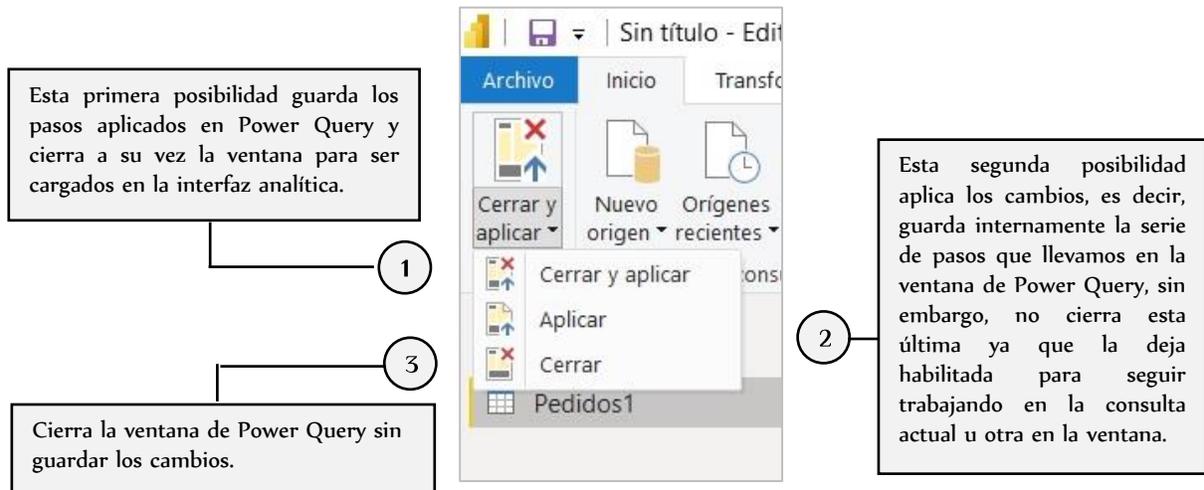


Figura 1. 32 – Opciones de Carga de Datos en Power BI Desktop: Cerrar y Aplicar, Aplicar y Cerrar

6. Nos iremos por la opción *Cerrar y Aplicar*.

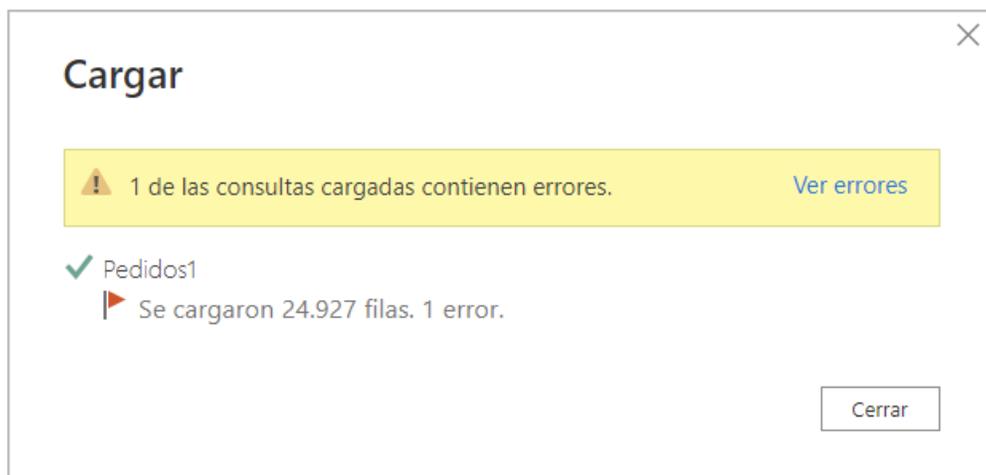


Figura 1. 33 – Cambios Aplicados en la Consulta y Advertencia de Error

Como se ve claramente arroja un error, no obstante, podemos seguir trabajando con normalidad, eso si es una sugerencia cortés de que hagamos tratamiento de estos antes de pasar a la parte de análisis, si queremos indagar de qué va el error puslamos clic en la etiqueta **Ver errores** en el cuadro de diálogo.

Ver
Errores

| Número de fila | Columna 1 | Columna 2 |
|----------------|-----------|-----------|
| 1 | Error | Error |

Figura 1. 34 – Consulta de Errores Generados

¡Riesgo: Advertencia!



Una buena consulta debe tener un manejo de excepciones, esto quiere decir que el propio query debe tomar ruta de acción si estos ocurren, garantizando que el proyecto no falle cada vez que se actualiza el reporte, si esto no se garantiza el sistema no es robusto.

Existen diversidad de técnicas y estrategias para manejar errores con *Power Query*, en el siguiente enlace podemos indagar sobre ello: <https://bit.ly/3w4ZpUf>

Por otra parte, es buena idea guardar el archivo de *Power BI*, este proceso es igual a guardar un archivo de *Excel*:

7. Vamos a la pestaña *Archivo* y ubicamos el comando *Guardar Como*, a continuación se despliega el cuadro de diálogo *Guardar Como* donde podemos navegar en nuestro computador para darle una localización deseada al archivo, así como un nombre y finalmente pulsar clic en el botón *Guardar*.

¡Riesgo: Advertencia!

Si un archivo de *Excel* está cifrado con contraseña, *Power BI* no lo podrá cargar, de hecho, arroja el siguiente error:



Si el tipo de protección utilizada es del tipo: *Proteger hojas*, estructura o modo lectura sí será posible cargar sus datos correctamente.

¡Posible Vía Alternativa!

Consiste en crear una carpeta de confianza, así se alojará el archivo allí y remover la contraseña, el *MVP Sergio Alejandro Campos* habla de ubicaciones de confianza en el siguiente vídeo: <https://www.youtube.com/watch?v=toN9LJuxhDg>.

Lectura incompleta de datos en Excel: *InferSheetDimensions*

Una situación que se presenta con “relativa” frecuencia con libros de *Excel*, es la lectura incompleta de un rango de datos no estructurado, para entenderlo utilizaremos el siguiente archivo:

¡Let's Play!



ARCHIVO DE TRABAJO: *En el compendio de archivos y carpeta Capítulo 1, puedes encontrar el archivo de Excel con nombre: CAPIEJ3-Distribuidora.xlsx.*



Recomendamos abrirlo desde su origen nativo (*Excel*) para darle un vistazo, percatarnos de su formato y ante todo visualizar el rango que contiene datos (A1:L32). La imagen a continuación lo presenta:

— ¿A qué se debe?

¡Riesgo: Advertencia!



Este error de lectura incompleta se debe a que cada libro de *Excel* en sus metadatos almacena las dimensiones de las hojas (según los rangos donde hay datos), resulta que *Power Query* por optimización lee estos metadatos para proceder con la vista previa y cargar, sin embargo, si estas dimensiones son incorrectas, porque algún sistema lo genera mal, entonces, *Power Query* los leerá inadecuadamente.

— ¿Cómo diagnosticar las dimensiones de una hoja de Excel?

1. Cambie el nombre del archivo .xlsx por una extensión .zip.
2. Abra el archivo en Explorador de archivos.

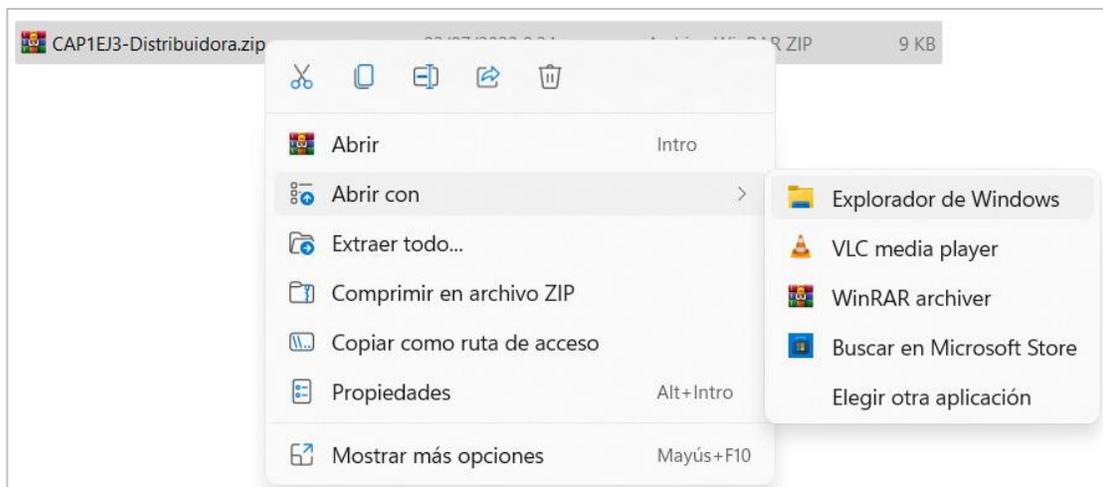


Figura 1. 37 – Navegar en .zip de Excel con el Explorador Windows

3. Vaya a xl\worksheets.
4. Copie el archivo xml de la hoja problemática (por ejemplo, *Sheet1.xml*) del archivo ZIP en otra ubicación.

| Nombre | Tipo | Tamaño comprimido | Protegido ... | Tamaño |
|------------|-----------------------|-------------------|---------------|--------|
| sheet1.xml | Archivo de origen XML | 2 KB | No | 13 KB |

Figura 1. 38 – Navegar en .zip de Excel con el Explorador Windows

5. Abrimos el `.xml` con un editor que consideremos apropiado, puede incluso ser el *bloc de notas*, para allí inspeccionar las primeras líneas del archivo (En la imagen más abajo se utiliza **Visual Studio Code** para abrir el archivo, además, se ha formateado el código con la **extensión XML de Red Hat**, para una lectura más fácil en el presente libro).
6. En el archivo buscamos la etiqueta `<dimension .../>`, para nuestro caso en la línea 4: `<dimension ref="A1:A34" />`.

NOTA

En algunos casos se da que lee más allá de donde hay datos, arrojando así una dimensión demasiado grande, por ejemplo, tener datos de A1 a F20 reales, pero *Power Query* lee de A1 a NY8000.

Lo anterior, genera problemas de rendimiento

```
C: > Miguel Caballero > EFB > C - Products Development > Libros > 09 - Trilogía de Power Query y Lenguaje M > Libro 1 - Power Q
1  <?xml version="1.0" encoding="UTF-8"?>
2  <worksheet xmlns="http://schemas.openxmlformats.org/spreadsheetml/2006/main" xmlns:mc="http
   markup-compatibility/2006" xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/r
   microsoft.com/office/spreadsheetml/2009/9/ac" mc:Ignorable="x14ac">
3    <sheetPr>
4      <outlinePr summaryBelow="0" />
5    </sheetPr>
6    <dimension ref="A1:A34" />
7    <sheetViews>
8      <sheetView workbookViewId="0" />
9    </sheetViews>
10   <sheetFormatPr defaultRowHeight="15" x14ac:dyDescent="0.25" />
11   <cols>
```

Figura 1. 39 – Metadatos de la Hoja 1 de un Archivo de Excel en Particular

- Si el archivo tiene un atributo de dimensión que apunta a una sola celda (como `<dimension ref="A1" />`), *Power Query* usa este atributo para buscar la fila inicial y la columna de los datos de la hoja.
- Sin embargo, si el archivo tiene un atributo de dimensión que apunta a varias celdas (como `<dimension ref="A1:A34" />`), *Power Query* usa este rango para buscar la fila y columnas iniciales, así como la fila y las columnas finales. Si este intervalo no contiene todos los datos de la hoja, algunos de los datos no se cargarán.

— ¿Cómo corregirlo?

Vamos al paso **Origen** en el panel de pasos aplicados, allí veremos la función del *Lenguaje M Excel.Workbook* que permite cargar los datos:

```
= Excel.Workbook ( File.Contents ( "C:\Dat" ), true, null )
```

Figura 1. 40 – Función Excel.Workbook generada por defecto en Power Query

El segundo argumento lo reemplazamos por:

```
[ InferSheetDimensions = true ]
```

Además eliminamos el tercer argumento, quedando así:

```
= Excel.Workbook ( File.Contents ( "C:\Dat" ), [ InferSheetDimensions = true ] )
```

Figura 1. 41 – Función Excel.Workbook modificada para Inferir Datos directamente de la Hoja

Ventajas y Desventajas

No está demás tener a la vista cuáles son las ventajas y desventajas del conector *Excel*.

Tabla 1. 1

Ventajas del Conector Excel

Ventajas

La forma más intuitiva de alimentar con datos a Power BI (Power Query)

Es un origen de datos universal, porque la mayoría de programas tienen como opción exportar a Excel

De una conector de Excel se pueden extraer: Tablas, Hojas, Nombres y Áreas de impresión

Tabla 1. 2

Desventajas del Conector Excel

Desventajas

Es una de las fuentes que más presenta datos por fuera del formato tabular

Excel no puede almacenar cantidades masivas de datos

Los archivos de Excel cifrados no pueden ser cargados

Las consultas no se pueden plegar

En el presente libro no es viable hacer el mismo ejercicio explicativo que hicimos con el conector de Excel con todos los demás orígenes, de hecho, sólo la categoría Archivo implicaría 8 conectores. Por ello, en este rubro dejamos que sea lector(a) quien inspeccione y familiarice con aquellos que más utiliza en su ámbito laboral, por ejemplo, si utiliza SQL Server se percatará de que consiste en proporcionar las credenciales:

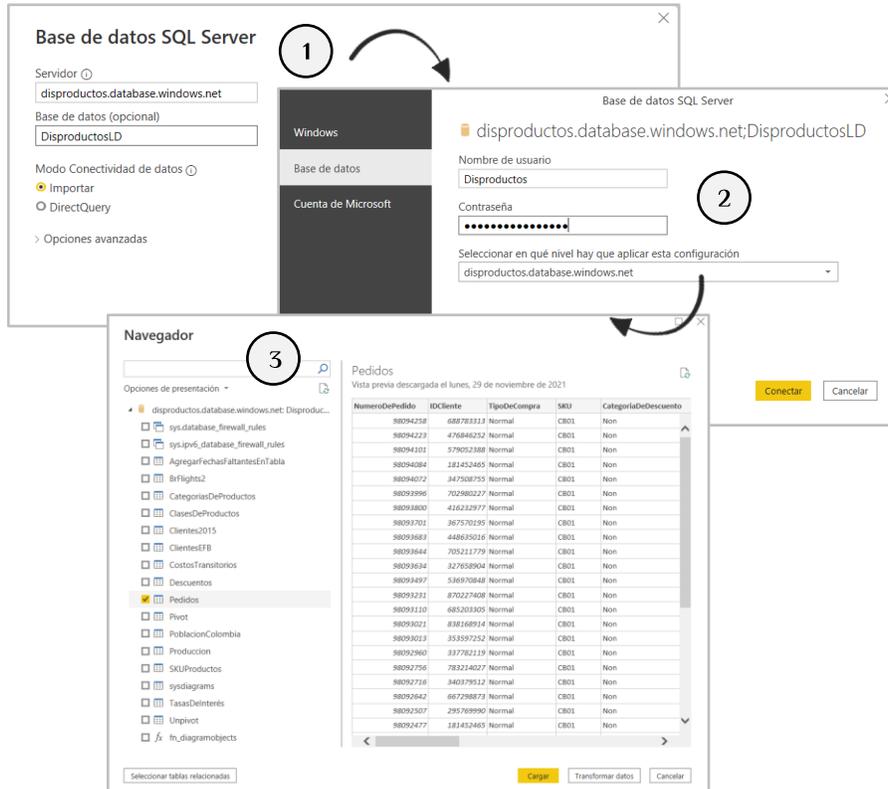


Figura 1. 42 – Credenciales Conector SQL Server en Power BI Desktop

– Además el firewall debe identificar la IP (consulta con tu área de sistemas).

¡Referencia Cruzada!



Un libro que hace un recorrido más exhaustivo por los distintos conectores es el título de Adam Aspin de *Power Query*, si bien no recorre todos, da un vistazo a bases relacionales como: *SQL Server* y *Oracle*, por el lado de la web explora: *Azure Blob Storage* y páginas web, además, brinda ejemplos de conexiones *ODBC* y *OLD BD*.

Data Mashup with Microsoft Excel Using Power Query and M
 ISBN: 978-1484260173 (<https://amzn.to/3tSAjFM>)

Power Query Extensiones



A estas alturas es bien sabido que *Power Query* trata de que los usuarios lleguen a las soluciones sin necesidad de código, y en esto hace un trabajo excepcional como notaremos en las páginas que yacen de aquí en adelante, no obstante, el mundo está lleno de retos, por esto en muchas ocasiones nos encontraremos con escenarios que no resuelven la interfaz.

— Por lo previo nace:



Zona de Power Query

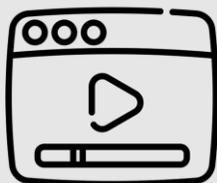
El sitio Web: <https://powerquery.zone/> es una página que recopila soluciones personalizadas para problemas ETL en *Power Query*, estas soluciones arrojan una función personalizada de fácil utilización para resolver problemas comunes, que no brinda por defecto el editor de *Power Query* o el *Lenguaje M*, algunos ejemplos:

- Dividir nombres y/o apellidos compuestos.
- Obtener una tabla con los días feriados (festivos) de tu país.
- Convertir abreviaturas de mes complejos o no estándar a números.
- Remover espacios intermedios excesivos entre palabras.
- Cálculo robusto de duración entre horas.
- Remover columnas vacías dinámicamente de forma óptima.
- Retornar celda o fila anterior.

NOTA

Power Query Zone es un proyecto del equipo de *Power Skill Tech*, el cual no sólo añade soluciones de sus miembros, sino también de la comunidad y de las inteligencias artificiales como *Chat GPT* y *Microsoft Copilot*. creadas con buen prompting y depuradas con conocimiento, para así contar con un repositorio lo más completo posible.

En el siguiente vídeo se presenta el proyecto:



<https://bit.ly/3MLMDQx>

— ¿Cómo se utiliza Power Query Zone?

Supongamos que se necesita fechas ferias o festivas para tu país en un intervalo de años determinados, en Power Query no existe una opción nativa para ello, sin embargo, en Power Query Zone contamos con la función personalizada.

**Tabla con
Feriados**

1. Mediante tu navegador de preferencia visitas la web: <https://powerquery.zone/>



Figura 1. 43 – Página web de Power Query Zone

2. En el extremo derecho, veremos la pestaña: Funciones, pulsamos clic allí para navegar por todas las funciones personalizadas disponibles.



Figura 1. 44 – Navegando por Power Query Zone

— Una vez localizada la función deseada:

3. Pulsa clic en “**Leer Más**” para navegar a la página dedicada de la función personalizada.
4. Desplazarse en la página hasta encontrar un recuadro rojo que dice: *Clic aquí para ver función*, esto es una opción en todas las páginas de las funciones personalizadas de *Power Query Zone*, pulsa clic en el recuadro.

Función Table.GenerateHoliDays

Para solucionar este caso, en una consulta en blanco en el editor de *Power Query* de *Power BI* o *Excel*, copie y pegue la función personalizada debajo, renómbrela como: [Table.GenerateHoliDays](#) (aunque el nombre puede ser cualquiera de tu preferencia), si deseas añadir la función a la librería de funciones M para quede 100% nativa, por consiguiente sea posible llamarla con el punto, es decir: [Table.GenerateHoliDays](#), mira los detalles de cómo hacerlo [en este vídeo](#).

▼ Clic Aquí para Ver Función

La función contempla los países de Iberoamérica y alimenta sus datos de la web: [cuandoenelmundo.com](#).

Figura 1. 45 – Recuadro con todo el código de la función personalizada

5. Nos trasladamos hasta la parte final del código, para observar en el extremo derecho una etiqueta que dice “*View Raw*”, pulsamos clic allí.

```

286 in
287     Value.ReplaceType (
288         Función,
289         Documentacion
290     )

```

Table.GenerateHoliDays hosted with ❤️ by GitHub view raw

Figura 1. 46 – “View Raw” en recuadro de código de función

6. En la página de *GitHub* con sólo el código, pulsamos la combinación de teclas *Ctrl + A* para seleccionar todo y luego *Ctrl + C* para copiar (es la forma que recomendamos para que no se escape nada del código).

3. En cualquier editor de *Power Query* (exceptuando el de *Power Automate*) se pulsa clic derecho en una parte en blanco del *Panel de consultas (Queries)*, para que en el menú que se despliega ubicar la opción: *Nueva consulta (New query)*:

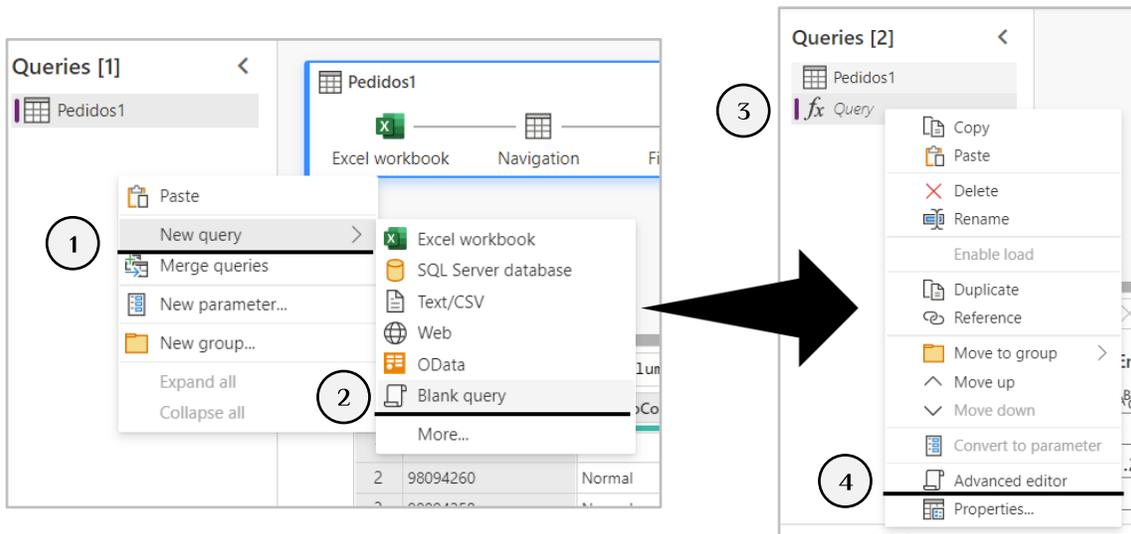


Figura 1. 47 – Nueva Consulta en Blanco

4. En la nueva consulta que aparece, pulsar clic derecho encima para elegir *Edior avanzado*, allí borrar todo el código y pegar el copiado de *Power Query Zone*, para finalmente pulsar clic en el botón *Listo*.

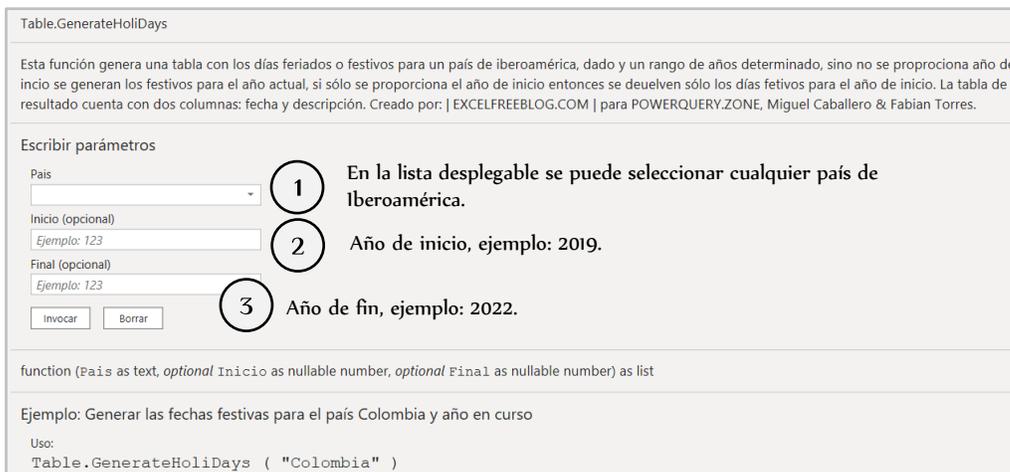


Figura 1. 48 – Función personalizada en Power Query Desktop

Pausemos un Momento



El complemento *Bravo*  para *Power BI* del equipo de *SQL BI* proporciona una forma rápida y fácil para obtener la tabla de feriados, sin embargo, a la fecha de escritura de este libro no están disponibles los países de América Latina (*proyectado a cambiar*), adicionalmente, si deseas obtener y trabajar con la tabla de feriados más allá de *Power BI Desktop* sigue siendo una buena opción la función personalizada de *Power Query Zone*.

Pausemos un Momento



Las funciones personalizadas se pueden volver 100% nativas en *Power BI Desktop*, para no tener que importarlas una y otra vez, en el siguiente vídeo puedes conocer el detalle:

- <https://www.youtube.com/watch?v=lQUqkeJT8lk&t=560s>

También es factible tener las funciones personalizadas siempre disponibles en *Power Query Online*, el siguiente artículo detalla cómo:

- <https://powerbisp.com/creacion-y-utilizacion-de-funciones-personalizadas-en-flujos-de-datos>

El Siguiete Paso

Este capítulo génesis ha sido un viaje de descubrimiento, una mezcla armoniosa de *teoría, práctica y utilidades generales*, similar a las notas iniciales de una sinfonía bien orquestada. Sin embargo, es momento de adentrarnos en la transformación y la resolución de *problemas ETL*. El próximo capítulo nos lleva a un nuevo nivel, iniciando con las operaciones que hemos denominado de *primer grado*. Aunque pueden parecer las más básicas, te invito a que te sumerjas en un estudio meticulado de ellas. Estará repleto de utilidades y desgloses totales de las operaciones fundamentales.