SALVADOR RAMOS

Excel 2013 Y Power Query

Recopila datos, olvídate de tareas tediosas y repetitivas.



Analiza tu Negocio con Excel y Power BI Aprende de tus datos

VOL. III

ADVERTENCIA LEGAL

Todos los derechos de esta obra están reservados a SolidQ[™] Press.

El editor prohíbe cualquier tipo de fijación, reproducción, transformación o distribución de esta obra, ya sea mediante venta, alquiler o cualquier otra forma de cesión o comunicación pública de la misma, total o parcialmente, por cualquier sistema o en cualquier soporte, ya sea por fotocopia, medio mecánico o electrónico, incluido el tratamiento informático de la misma, en cualquier lugar del mundo.

La vulneración de cualquiera de estos derechos podrá ser considerada como una actividad penal tipificada en los artículos 270 y siguientes del Código Penal.

La protección de esta obra se extiende al mundo entero, de acuerdo a las leyes y convenios internacionales.

© SolidQTM Press, 2015

Título:

Excel 2013 y Power Query.

Recopila datos, olvídate de tareas tediosas y repetitivas

Autor: Salvador Ramos

ISBN: 978-84-940719-3-5

Serie:

Analiza tu Negocio con Excel y Power BI.

Aprende de tus datos

SolidQ Global S.A. Apartado de correos 202 03340 Albatera, Alicante, España http://www.solidq.com

AUTOR



Soy **experto en BI & Analytics**Director de Formación en SolidQ
SOL Server MVP desde el año 2003

¿En qué te puedo ayudar?

Quiero acompañarte en tu transición hacia el mundo del BI & Analytics y Big Data. Tanto si estás dando un giro a tu carrera profesional, como si buscas una mayor especialización en estas áreas.

Estoy especializado en Bases de datos, Data Warehousing, ETL, técnicas de Visualización y en el uso de tecnologías Microsoft (SQL Server, Integration Services, Analysis Services, Reporting Services, Performance Point Services, Self-service BI, Excel, Power BI, Microsoft Azure).

Te invito a que conozcas mi <u>blog personal</u> (pincha <u>aquí</u>), a que te <u>suscribas a él</u> para recibir tu regalo (pincha <u>aquí</u>) y estés al tanto de todas las novedades (*libros y artículos que publico, conferencias en las que participo, cursos y masters que imparto, material que regalo* ...).

Por confiar en mí, te ofrezco gratis mi libro gratuito "Microsoft Business Intelligence: Vea el cubo medio lleno". Descárgalo <u>aquí</u>.

Sígueme también en: http://www.salvador-ramos.com Linkedin: http://www.linkedin.com/in/SalvadorRamos

Twitter: @salvador ramos

SOLIDQ



Think Big. Move Fast.

SolidQ, desde el año 2002, suministra servicios para plataformas Microsoft que le ayudan a diseñar, integrar y optimizar su utilización de datos.

Combina una amplia experiencia técnica y de implementación en el mundo real, con un compromiso firme en la transferencia de conocimiento, dada la combinación única de dotes lectivas y experiencia profesional que nuestros mentores ofrecen. De este modo, no solamente ayudamos a nuestros clientes a solventar sus necesidades tecnológicas, sino que somos capaces de incrementar la capacidad técnica de sus profesionales, dándoles una ventaja competitiva en el mercado. Por eso llamamos **Mentores** a nuestros expertos: por su compromiso en asegurar el éxito de su empresa y de sus equipos profesionales a largo plazo.

Nuestros expertos son profesionales reconocidos en el mercado, con más de 100 premios Microsoft MVP (Most Valuable Professional) obtenidos hasta la fecha. Se trata de autores y ponentes en las conferencias más importantes del sector, con varios centenares de ponencias presentadas en conferencias nacionales e internacionales durante los últimos años.

Nuestra misión es la de **transmitir todo el conocimiento adquirido** resolviendo problemas del mundo real para miles de clientes, escribiendo artículos y libros, publicando whitepapers, creando contenidos educativos y formando a decenas de miles de trabajadores de TI en todo el mundo, para que los proyectos de nuestros clientes obtengan los mayores éxitos. Esta transferencia de conocimiento la realizamos fundamentalmente con dos tipos de servicios:

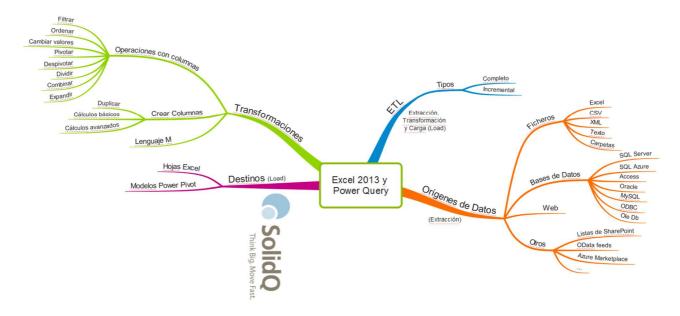
Consultoría: hazlo bien la primera vez (haz clic aquí)

Mentoring: conoce tu potencial personal y mejora tus decisiones (haz clic aquí)

Formación: la mejor inversión posible es pagar por el conocimiento de otros. Conoce nuestro **Plan Formativo** (haz clic <u>aquí</u>) y nuestro **Calendario** (haz clic <u>aquí</u>)

Publicaciones: ponemos nuestros conocimientos a su alcance. Acceda a nuestros **blogs** (haz clic <u>aquí</u>) y a nuestras **publicaciones**, la mayoría gratuitas (haz clic <u>aquí</u>)

MAPA MENTAL



ÍNDICE

ADVERTENCIA LEGAL	0
AUTOR	2
SOLIDQ	3
MAPA MENTAL	4
Introducción	6
Trabajando con Power Query	8
Orígenes de datos	10
Transformaciones	22
Menu 'Home'	23
Menu 'Transform'	26
Menu 'Add Column'	29
Menu 'View'	30
Destinos	35
Lenguaje M	38
Power Query vs SSIS (SQL Server Integration Services)	38
Conclusiones	39
ENLACES ESENCIALES	40
GRACIAS	41

Introducción

Una vez que tenemos identificados los orígenes de datos y definido el Data Warehouse y/o los Data Marts, básicamente lo que tenemos es una estructura preparada para el análisis de la información, pero hasta el momento, vacía.

Los **procesos ETL** (por sus siglas en inglés: Extract, Transform and Load) son los que se encargan de poblar de información, pasando una serie de procesos a partir de la extracción desde los diversos orígenes, que hagan las transformaciones y cálculos necesarios, que hagan tareas de limpieza, que detecten incoherencias, datos sin completar, validaciones de reglas de negocio y demás. Y una vez hechas todas estas integraciones y transformaciones para conseguir la estructura y calidad del dato necesaria para almacenarla en el destino, que la graben (carguen) en el Data Warehouse y/o en los Data Marts.

Los procesos ETL se ejecutan periódicamente (cada minuto, cada hora, cada día, cada semana, cada mes, cada semana de lunes a viernes, etc.), siendo a día de hoy la ejecución diaria la más habitual (normalmente, durante el periodo de poca o ninguna actividad en los sistemas transaccionales, que suele ser por las noches), para que así, cuando los usuarios de negocio lleguen cada mañana, tengan toda la información hasta el cierre del día anterior actualizada.

Diseñar un buen proceso ETL es vital para cualquier proyecto. Tanto por la calidad del dato, como por duración de su ejecución.

Hoy en día la mayoría de las herramientas que existen para realizar los procesos ETL están orientadas a su instalación en servidores y a su uso por parte del departamento de TI.

En el caso de la plataforma Microsoft, la herramienta es SQL Server Integration Services (SSIS). Esta es la definición que hace Microsoft de ella:

"Microsoft Integration Services es una plataforma para la creación de soluciones empresariales de transformaciones de datos e integración de datos. Integration Services sirve para resolver complejos problemas empresariales mediante la copia o descarga de archivos, el envío de mensajes de correo electrónico como respuesta a eventos, la actualización de almacenamientos de datos, la limpieza y minería de datos, y la administración de objetos y datos de SQL Server. Los paquetes pueden funcionar por separado o conjuntamente con otros paquetes para hacer frente a las complejas necesidades de la empresa. Integration Services puede extraer y transformar datos de muchos orígenes distintos, como archivos de datos XML, archivos planos y orígenes de datos relacionales, y, posteriormente, cargarlos en uno o varios destinos.

Integration Services contiene un variado conjunto de tareas y transformaciones integradas, herramientas para la creación de paquetes y el servicio Integration Services para ejecutar y administrar los paquetes. Las herramientas gráficas de Integration Services se pueden usar para crear soluciones sin escribir una sola línea de código. También se puede programar el amplio modelo de objetos de Integration Services para crear paquetes mediante programación y codificar tareas personalizadas y otros objetos de paquete."

Si desea profundizar en sus conocimientos sobre SQL Server Integration Services (SSIS) puede descargar el libro gratuito "**Microsoft BI: Vea el cubo medio lleno**", autor: Salvador Ramos, editorial SolidQ Press. http://bit.ly/LibroBlcubo y también puede compartir este enlace con los técnicos de su equipo.

Por el momento sólo queremos que se quede con la idea de que Microsoft ofrece una potente herramienta ETL orienta al uso por el departamento de TI. Esto ocurre igual en las plataformas de BI de otros fabricantes; la diferencia es que Microsoft ha creado una herramienta ETL adicional, orientada al usuario de negocio, llamada **Power Query**, que *se integra en Excel 2013* como un "complemento" (*add-in*) que se puede descargar de forma gratuita desde su web, y una vez instalada, simplemente, agrega una opción de menú, también llamada *Power Query*, a la "cinta" (*ribbon*) de menús de Excel. Evidentemente no tiene la potencia y flexibilidad de una herramienta ETL para BI corporativo y orientada al personal de TI, pero es ideal para soluciones de BI Personal, tanto por su potencia como por su facilidad de uso. A continuación estudiaremos más a fondo esta herramienta.

Si está habituado a construir informes y cuadros de mando a base de copiar y pegar datos de los orígenes a sus hojas Excel, y a partir de ahí seguir la misma secuencia de pasos manuales, una y otra vez, hasta convertirlos en información útil y aplicar complejas fórmulas, que tiene que revisar todo cada vez que cambian los datos, por si han cambiado los rangos de celdas u otros detalles. **Power Query** le ayudará a automatizar, de una vez por todas, estos procesos y se complementará con el resto de componentes de Power BI para facilitarle el análisis de información.

<u>Nota</u>: También hay una versión de Power Query para Excel 2010 que puede descargar e instalar de forma gratuita. Son muy similares ambas versiones, aunque aquí vamos a utilizar Power Query para Excel 2013.

Trabajando con Power Query

Power Query es un complemento para Excel que nos permite conectarnos a una gran diversidad de orígenes de datos (Access, MySQL, SQL Server, Oracle, archivos de texto, fuentes públicas de datos, OData, Marketplace, etc.) y transformar los datos obtenidos en función de nuestras necesidades. Puede crear o eliminar columnas, renombrarlas, combinar tablas, unir tablas, hacer cálculos; en definitiva partir de los datos de origen y hacer todas las transformaciones necesarias para convertirlos en información preparada para el análisis, y finalmente almacenarlos, tanto en una hoja Excel tradicional (*WorkSheet*) como en los nuevos modelos de **Power Pivot** (*Excel Data Model*). Todos estos procesos quedan almacenados para su posterior ejecución que refresque los datos, tan sólo pulsando el botón "Refrescar", sin necesidad de repetir una y otra vez esas tareas repetitivas y tediosas de transformación de forma manual que realizamos con el Excel tradicional cada vez que queremos actualizar la información.

En otras palabras, lo podríamos definir como una "herramienta ETL (*Extract, Transform and Load*) orientada al usuario final" (no es necesario saber de programación) y que está totalmente integrada en Excel. Como puede comprobar en la siguiente imagen, su interfaz es un menú más en la "cinta" (*Ribbon*) de Excel.

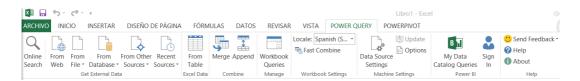


Figura 0-1 Power Query Ribbon en Excel

<u>Nota</u>: Aunque disponga de Excel en Español, las opciones de Power Query no han sido traducidas, sólo están disponibles inglés. Puede apreciarlo en la imagen anterior.

En la siguiente imagen vamos a ubicar **Power Query** junto con el resto de componentes de Power BI para Excel, con el objetivo de tener una visión global:

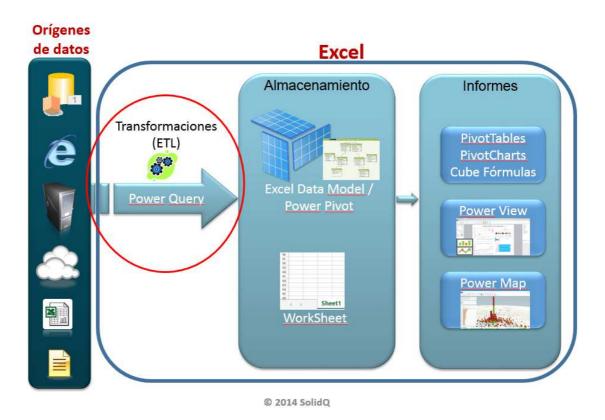


Figura 0-2 Componentes de Excel Power BI. Power Query.

Orígenes de datos

Power Query es capaz de conectarse a una gran variedad de orígenes de datos, esta es una de sus características más destacadas. Puede conectarse a múltiples formatos de archivos planos, a archivos XML, a páginas web, a la mayoría de las bases de datos del mercado, a listas de SharePoint, incluso a Facebook y fuentes Open Data. Más adelante estudiaremos con detalles las conexiones más significativas.

A continuación mostramos la parte de la "cinta" (*Ribbon*) llamada "*Get External Data*" donde tenemos disponibles todas las opciones que nos permiten obtener datos externos:

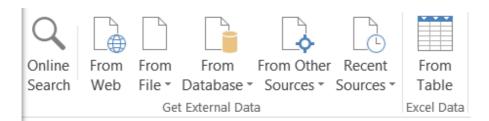


Figura 0-3 Opciones para Obtener Datos Externos (Get External Data).

A continuación vamos a ir estudiando cada una de estas opciones. Vamos a comenzar por las opciones relacionadas con datos que se obtienen desde páginas web:

- Online Search: permite realizar búsquedas de términos que devuelven páginas web donde hay información relacionada con ellos (la mayor parte de ellas provienen de Wikipedia).
- From Web: permite introducir una URL (dirección web) y obtiene de ella las diversas tablas y datos susceptibles de ser extraídos, para que usted pueda previsualizarlos, y elegir los que quiere obtener.
 Puede hacer una prueba pulsando el botón "Online Search", introduciendo la siguiente URL https://es.finance.yahoo.com/q/cp?s=%5EIBEX y previsualizando el resultado de 'Table1' en el panel que aparece a la derecha de su pantalla (ventana "Navigator").

From From Other Recent
Database Sources Sources
From Excel
Import data from a
Microsoft Excel workbook.

From CSV
Import data from a
comma-separated value file.

From XML
Import data from an XML
file.

From Text
Import data from a text file.

From Folder
Import metadata and links

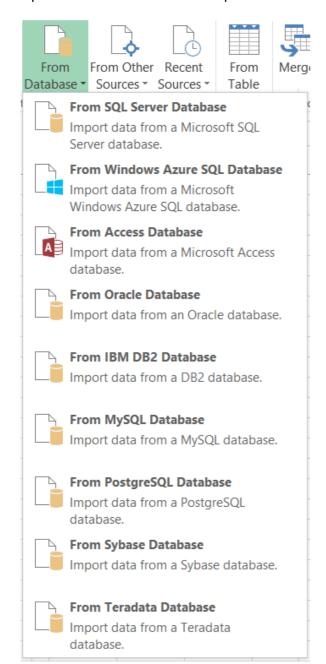
Opciones relacionadas con la obtención de datos a partir de ficheros:

Figura 0-4 Opciones 'Desde Fichero' (From File).

about files in a folder.

- From Excel: importa datos contenidos en libros Excel. En la ventana
 "Navigator", que aparece en la parte derecha nos mostrará los diversos
 grupos de datos y tablas que ha detectado dentro de ese libro Excel y
 podremos previsualizarlos, para posteriormente obtener los que vayamos
 seleccionando.
- From CSV: CSV es un formato estándar que se utiliza con mucha frecuencia. En él los datos vienen separados por comas (o a veces por punto y coma), por ejemplo, lo utilizan algunos bancos para entregarnos información de nuestras cuentas o de movimientos realizados. Si el fichero seleccionado cumple los estándares CSV, lo previsualiza directamente e inicia el proceso transformaciones y carga de él.
- **From XML**: importa datos desde archivos XML y permite navegar por las diferentes estructuras de datos de dicho archivo, seleccionando las columnas de las que queremos obtener datos.
- **From Text**: permite obtener datos de cualquier archivo de texto, siempre que en éste detecte una determinada estructura que se pueda convertir en una

- tabla. Habitualmente son ficheros que llevan delimitadores de columnas (coma, punto y coma, tabuladores, etc.) o que cada columna tiene una longitud fija.
- **From Folder**: puede obtener información sobre las características, propiedades, archivos y carpetas que hay dentro de la carpeta indicada. Por ejemplo podríamos obtener una tabla con una lista de características de los archivos y carpetas que hay en nuestro disco duro o en cualquier carpeta.



Opciones relacionadas con la importación desde Bases de Datos:

Figura 0-5 Opciones 'Desde Base de datos' (From Database).

- **From SQL Server Database**: permite obtener datos almacenados en bases de datos de SQL Server.
- **From Windows Azure SQL Database**: permite obtener datos almacenados en bases de datos de Azure SQL, que tenemos en la nube de Microsoft.

- From Access Database: permite obtener datos almacenados en bases de datos Access.
- **From Oracle Database**: permite obtener datos almacenados en bases de datos Oracle.
- **From IBM DB2 Database**: permite obtener datos almacenados en bases de datos IBM DB2.
- **From MySQL Database**: permite obtener datos almacenados en bases de datos MySQL.
- **From PostgreSQL Database**: permite obtener datos almacenados en bases de datos PostgreSQL.
- **From Teradata Database**: permite obtener datos almacenados en bases de datos Teradata.

Según el sistema gestor de bases de datos al que nos conectemos le pedirá una serie de características de acceso, que habitualmente nos tendrá que facilitar el departamento de TI.

Veamos un ejemplo de acceso a la base de datos SQL Server llamada 'TBI01_Origen' que se encuentra en el servidor 'pc-salva':



Figura 0-6 Conexión a SQL Server.

Opciones relacionadas con la importación desde otras fuentes:

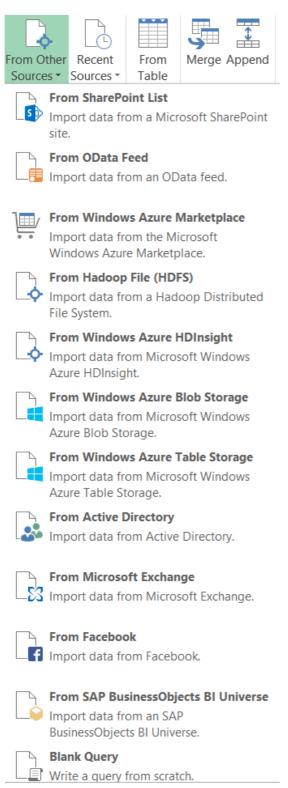


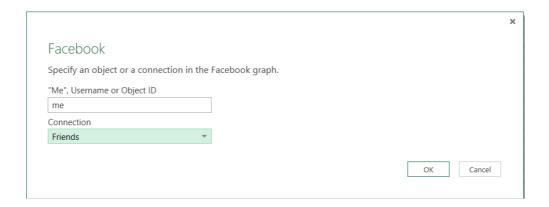
Figura 0-7 Conexión a Otras fuentes de datos.

- **From SharePoint List**: permite obtener datos almacenados en una lista de SharePoint.
- **From OData Feed**: permite obtener datos almacenados en fuentes de datos Open Data.
- **From Windows Azure Marketplace**: permite obtener datos almacenados en Azure Marketplace.
- **From Hadoop File (HDFS)**: permite obtener datos almacenados en Hadoop (Big Data).
- From Windows Azure HDInsight: permite obtener datos almacenados en HDInsight.
- **From Windows Azure Blob Storage**: permite obtener datos almacenados en Azure Blob Storage.
- **From Windows Azure Table Storage**: permite obtener datos almacenados en Azure Table Storage.
- **From Active Directory**: permite obtener datos almacenados en el Directorio Activo de la empresa.
- **From Microsoft Exchange**: permite obtener datos almacenados en Microsoft Exchange (servidor de correo).
- From Facebook: Permite obtener datos almacenados en Facebook, que estén accesible desde la cuenta que utilicemos para conectarnos. Por ejemplo podemos extraer los amigos asociados dicha cuenta, información de publicaciones, etc.
- **From SAP BusinessObjects BI Universe**: permite obtener datos almacenados en los universos de SAP Business Objects.
- **Blank Query**: finalmente, podemos partir de una consulta en blanco e ir añadiendo los elementos que estimemos oportunos.

No se preocupe si no conoce alguna de estas fuentes o no tiene claro cómo realizar la conexión a ellos, es algo habitual y que en ciertas ocasiones necesita la colaboración del departamento de TI para que nos den las indicaciones oportunas. Pero en otros casos no es así, simplemente guiándonos por las pantallas en las que nos van pidiendo los datos, podremos hacerlo nosotros mismos.

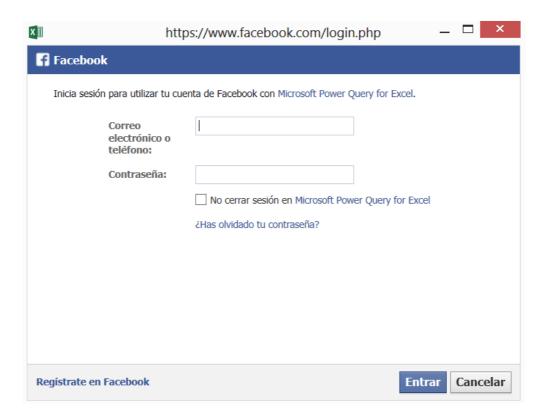
Veamos a continuación cómo conectarnos a una cuenta de Facebook, para que quede claro que es algo sencillo de realizar por nosotros mismos. Para ello seguiremos los siguientes pasos:

1. Elegimos la opción 'From Other Sources' y allí hacemos clic en 'From Facebook', apareciendo la siguiente ventana de conexión.



Donde seleccionamos los valores que se ven en la imagen anterior. Vamos a obtener la lista de "Amigos" de nuestra cuenta de Facebook. Si pincha en el desplegable que hay debajo de la palabra 'Connection' podrá ver las diversas opciones para la obtención de información de Facebook que hay disponibles (Actividades, Comentarios, Amigos, Feeds, Me gusta, Enlaces, Posts, etc.).

- 2. Pulsamos el botón 'OK'
- 3. Nos aparece una pantalla pidiéndonos las credenciales, donde pulsaremos en el botón 'Sign in' e introduciremos nuestro usuario y contraseña de Facebook.



4. Pulsaremos el botón 'Entrar'.

- 5. Nos informa del usuario con el que estamos conectados a Facebook y pulsamos el botón 'Save'.
- 6. Nos aparece la ventana con los datos obtenidos, la misma a la que llegamos una vez obtenemos acceso a la estructura de datos elegida, independientemente del origen seleccionado, la venta del 'Query Editor'. A partir de ahí, como se hace con cualquier otro origen de datos, podemos aplicar las transformaciones que deseemos y llevarlos a nuestro destino. Esto lo estudiaremos en los siguientes apartados. De momento nos quedamos con que hemos conseguido ver los datos tal cual están en el origen para la selección realizada.

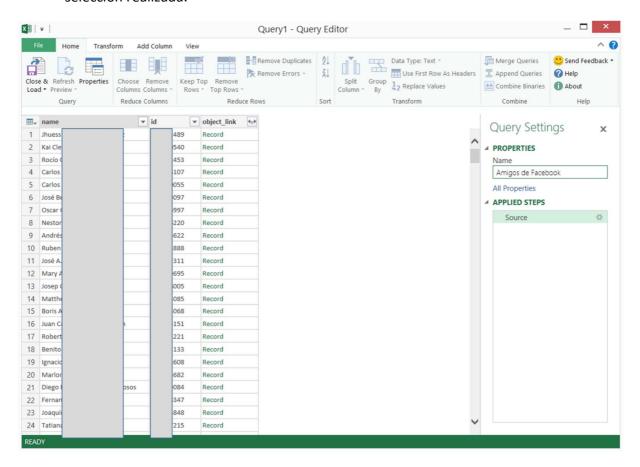


Figura 0-8 Ventana del 'Query Editor'.

El proceso para conectarnos a cualquier otra fuente de datos es muy similar, lo que irá cambiando son las pantallas donde nos va pidiendo los datos de conexión, que serán diferentes en cada caso, pero bastante intuitivas y claras.

También tenemos disponible la opción de la "cinta" (*Ribbon*) llamada "*Excel Data*" donde tenemos una sola opción:

- **From Table**: esta opción nos permite obtener datos que están dentro del propio libro Excel que estamos utilizando en este momento.

Como puede comprobar, hay acceso prácticamente a cualquier origen de datos de los que se utilizan hoy en día. Esta conectividad tan diversa es una de las fortalezas más destacadas de *Power Query*.

Una vez que hemos seleccionado el origen de datos, bien nos lleva directamente al editor, bien para hacerlo debemos seleccionar entre las diferentes estructuras de datos disponibles y a continuación, tenemos dos opciones mediante los botones que hay en la parte inferior:

- 'Edit': que nos lleva a la ventana del 'Query Editor' (la misma a la que hemos llegado en el ejemplo de obtención de la tabla con nuestros amigos de Facebook), donde podemos ir incluyendo las transformaciones que veamos oportunas y viendo directamente cómo van quedando los datos al aplicarlas. Más adelante lo veremos con detalle y estudiaremos las opciones de transformación disponibles.
- 'Load': que hace la carga directamente, sin permitirnos incorporar transformaciones adicionales. Nos ofrece la opción de cargar en una hoja Excel y/o en el modelo Power Pivot.

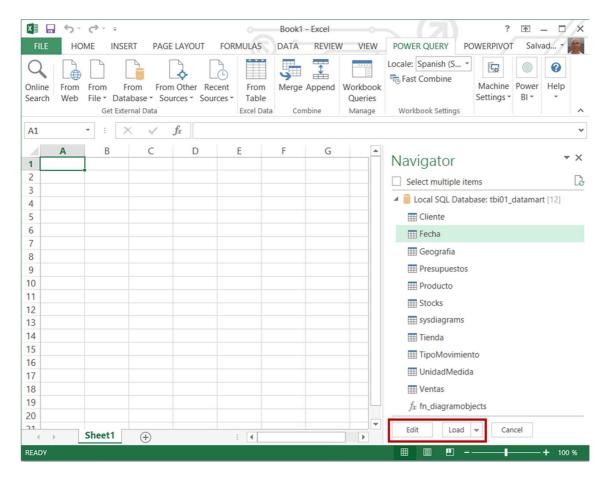


Figura 0-9 Ventana "Navigator", botones "Edit" y "Load".

Para finalizar, a modo de sintetizar los muchos elementos que hemos visto, exponemos una tabla resumen de los orígenes de datos disponibles:

Bases de datos	Ficheros	Internet	Otros
SQL Server	Excel	Web	Lista SharePoint
SQL Server Azure	CSV	OData	Azure Martketplace
Access	XML	Facebook	Hadoop
Oracle	Texto	Servicios Web	Azure HDInsight
IBM DB2	Carpeta		Azure Blob Storage
MySQL			Azure Table Storage
PostgreSQL			Active Directori
Sybase			Microsoft Exchange
Teradata			SAP Business Object

Transformaciones

Una vez que hemos terminado la selección de los datos que queremos obtener del origen, siempre llegamos a un punto común, la ventana 'Query Editor' que es la que nos permite hacer las transformaciones que estimemos necesarias a partir de los datos seleccionados. Veamos a continuación una imagen completa de esta ventana con el resultado de los datos obtenidos en el ejemplo de los amigos de la cuenta de Facebook:

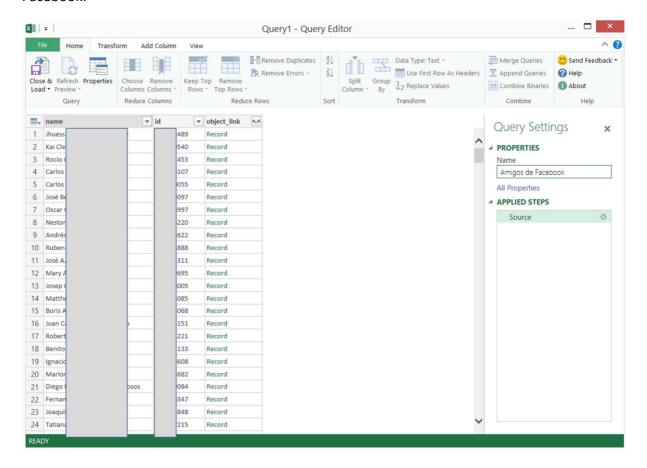


Figura 0-10 Ventana del 'Query Editor'.

Como hemos comentado anteriormente, Power Query ofrece una gran variedad de transformaciones que nos permiten ir convirtiendo los datos y haciendo los cálculos necesarios para realizar la integración entre diversas fuentes y obtener la estructura que hemos diseñado como destino aplicando las técnicas de modelado dimensional.

Vamos a pasar a continuación a conocer todas las transformaciones disponibles, haciendo un estudio de los tres elementos de menús disponibles en la "cinta" (*Ribbon*), que son: 'Home', 'Transform', 'Add Column' y 'View'.

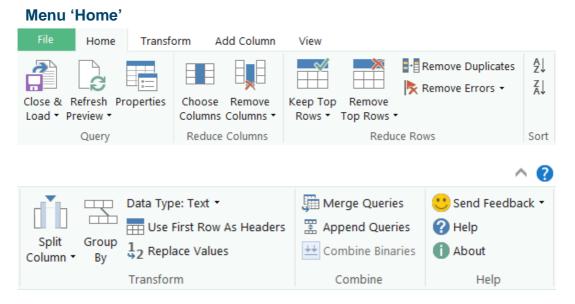


Figura 0-11 Ribbon, menú 'Home'.

Grupo de opciones '**Query**': son opciones relativas a la consulta al origen de datos.

- Close & Load: permite cerrar el proceso de extracción y transformaciones y proceder a la carga de los datos en el destino. Se verá con detalle más adelante, cuando se estudien los destinos disponibles.
- **Refresh Preview**: actualiza los datos del modelo. Elimina totalmente el contenido del modelo y lo vuelve a extraer del origen, aplicar los pasos con las transformaciones definidas y cargar en el destino.
- Properties: permite modificar el nombre y la descripción de la consulta actual.

Grupo de opciones '*Reduce columns*': permiten quitar columnas que no necesitamos.

- Choose Columns: permite seleccionar las columnas que queremos tener disponibles y eliminar las demás. Se pueden marcar o desmarcar cualquiera de las columnas de la consulta, quedando disponibles sólo las que dejemos marcadas.
- **Remove Columns**: permite eliminar columnas de la tabla.

Grupo de opciones 'Reduce Rows': permiten eliminar filas que no necesitamos.

- **Keep Top Rows**: permiten mantener un conjunto de filas, bien del principio, bien del final, bien un rango, eliminando las que queden fuera de la selección.

- **Remove Top Rows**: permiten eliminar un conjunto de filas, bien del principio, bien del final, bien un rango, manteniendo las que queden fuera de la selección. Es la operación contraria a la anterior.
- **Remove Duplicates**: elimina filas duplicadas, basándose en las columnas seleccionadas en ese momento.
- **Remove Errors**: permite eliminar o mantener las filas que contengan errores.

Grupo de opciones 'Sort': permiten realizar ordenaciones de los datos.

- **A-Z**: ordenación ascendente de las filas en base a la columna o columnas seleccionadas.
- **Z-A**: ordenación descendente de las filas en base a la columna o columnas seleccionadas.

Grupo de opciones '*Transform*': permiten realizar transformaciones de datos.

- **Split Column**: podemos dividir una columna en varias, bien por un delimitador, bien por una longitud fija. Por ejemplo, si tenemos una columna con el nombre y apellidos de los clientes, con las filas almacenadas de la siguiente forma: 'apellido1 apellido2, nombre', podríamos dividirla en dos columnas, una con el nombre y otra con los apellidos, indicando que el delimitador es la coma.
- **Group By**: podemos agrupar conjuntos de filas y aplicar funciones de agregación sobre ellos (suma, promedio, contar, mínimo, máximo, etc.). Por ejemplo, a partir de la tabla de clientes, podríamos agrupar por provincia y obtener una fila por cada provincia con el número de clientes que tenemos en ella
- **Data Type**: permite hacer conversiones de tipos de datos.
- **Use First Row AS Headers**: indica si la primera fila obtenida del origen son datos o son los encabezados de columnas. Si se pulsa, considera la primera fila como encabezados de columnas. Se suele usar en archivos de texto.
- **Replace Values**: es como el buscar y reemplazar de Excel, indicamos el valor a buscar y el valor por el que lo queremos reemplazar.

© SolidQ 2015 www.solidq.com pág. 24

Grupo de opciones 'Combine': permiten realizar combinaciones de tablas con datos.

- Merge Queries: la operación Merge implica unir columnas de dos tablas que tienen valores en común por otra columna, habitualmente se unen por las claves. El número de filas nunca será superior al número de filas de la tabla que más filas tenga. Esta imagen le será de ayuda para entender la operación:



- Append Queries: la operación Append implica unir dos tablas con la misma estructura en una sola, obteniendo una nueva tabla con las mismas columnas y con el total de filas de ambas tablas. Por ejemplo, si tenemos una tabla con las ventas de enero y otra con las ventas de febrero, podemos obtener una sola tabla con las ventas de enero y febrero. Esta imagen le será de ayuda para entender la operación:

Fecha	Cliente	Producto	Cantidad			Fecha	Cliente	Producto	Cantidad
01/01/2014	3	330	1			01/02/2014	1	470	1
02/01/2014	2	330	1			02/02/2014	4	330	1
03/01/2014	5	470	3			03/02/2014	7	330	3
04/01/2014	1	470	1			04/02/2014	2	488	1
29/01/2014	4	330	2			26/02/2014	5	573	2
30/01/2014	7	573	1			27/02/2014	1	470	1
31/01/2014	3	488	3			28/02/2014	4	330	3
		\					/		
			Fecha	Cliente	Producto	Cantidad			
			01/01/2014	3	330	1			
			02/01/2014	2		1			
			03/01/2014	5		3			
			04/01/2014	1	470	1			
				_		_			
			29/01/2014	4		2			
			30/01/2014	7		1	_/_		
			31/01/2014 01/02/2014	3		3			
			02/02/2014	1 4			_		
			03/02/2014	7		1			
			04/02/2014	2		1	↓		
			04/02/2014	2	400	1			
			26/02/2014	5	573	2			
			27/02/2014	1		1			
			28/02/2014	4		3			
			20/02/2014	4	330	3			

Menu 'Transform' Home Transform Add Column 🚅 Transpose Data Type: Text * **Expand** Reverse Rows 12 Replace Values 📮 Pivot Column Aggregate Aggregate Group Use First Row 🛵 Replace Errors 🖙 Unpivot Columns 🔻 🔣 Move 🔻 Count Rows As Headers Table Any Column Duration : Text Column Number Column Date & Time Column

Figura 0-12 Ribbon, menú 'Transform'.

Grupo de opciones '*Table'*: permiten realizar operaciones sobre el conjunto de la tabla. Algunas opciones también están en otros menús y se repiten aquí por usabilidad, evitando, en la medida de lo posible, que tengamos que estar cambiando de menús.

- Group By: podemos agrupar conjuntos de filas y aplicar funciones de agregación sobre ellos (suma, promedio, contar, mínimo, máximo, etc.). Por ejemplo, a partir de la tabla de clientes, podríamos agrupar por provincia y obtener una fila por cada provincia con el número de clientes que tenemos en ella. Está también en el menú 'Home'.
- **Use First Row AS Headers**: indica si la primera fila obtenida del origen son datos o son los encabezados de columnas. Si se pulsa, considera la primera fila como encabezados de columnas. Se suele usar en archivos de texto. Está también en el menú *'Home'*.
- **Transpose**: pasa la información de filas a columnas y viceversa. Tenga en cuenta que puede devolver un gran número de columnas si hay muchas filas, y que además perdería los nombres de columnas originales, quedando nombradas como Column1 a ColumnN.
- Reverse Rows: invierte el orden de las filas. Aunque existe esta opción, es recomendable utilizar las opciones de ordenación y conseguir con ellas la ordenación más adecuada.
- Count Rows: elimina las filas de la tabla y devuelve el número de filas de dicha tabla. Su única utilidad es almacenar el número de filas resultante, sin ningún otro dato adicional.

Grupo de opciones 'Any Column': permiten realizar operaciones sobre cualquier columna disponible en ese punto del proceso de transformación.

- **Data Type**: permite cambiar el tipo de datos de una columna. Está también en el menú 'Home'.
- Replace Values: es como el buscar y reemplazar de Excel, indicamos el valor a buscar y el valor por el que lo queremos reemplazar. Está también en el menú 'Home'.
- Replace Errors: reemplaza los errores por el valor que le indiquemos.
- **Fill**: permite rellenar valores nulos (*Null*) aplicando el valor de la fila anterior (*Fill Down*) o el de la fila posterior (*Fill Up*)
- **Pivot Column**: permite pivotar columnas de una tabla.
- **Unpivot Columns**: permite despivotar columnas de una tabla. Las operaciones *Pivot y Unpivot* cambian filas por columnas y viceversa, funciona de la misma forma que la operación 'Transponer' de Excel.
- **Expand**: equivale a pulsar el botón estudiado anteriormente y utilizado en uno de los ejemplos.
- **Aggregate**: sólo está activo en las columnas que contengan el botón (tablas relacionadas) y permite obtener datos agregados de las columnas numéricas que contienen. Por ejemplo, en la tabla 'Tienda' podemos obtener la suma de unidades de cada tienda desde la tabla **MovimientosTiendas** .
- **Move**: permite mover las columnas hacia la izquierda o derecha. Es igual que si pinchamos sobre una columna y la arrastramos a la posición deseada, resultando esta segunda opción más cómoda.

Grupo de opciones 'Text Column': permiten realizar operaciones específicas sobre las columnas que contienen texto (cadenas de caracteres).

- **Split Column**: podemos dividir una columna en varias, bien por un delimitador, bien por una longitud fija. Por ejemplo, si tenemos una columna con el nombre y apellidos de los clientes, con las filas almacenadas de la siguiente forma: 'apellido1 apellido2, nombre', podríamos dividirla en dos columnas, una con el nombre y otra con los apellidos, indicando que el delimitador es la coma. Está también en el menú '*Home*'.
- **Format**: permite, sobre la columna elegida: convertirla a mayúsculas o minúsculas, poner la primera letra de cada palabra en mayúsculas, eliminar espacios por la derecha e izquierda, y eliminar caracteres no imprimibles.
- Merge Columns: concatena las columnas seleccionadas en una sola columna, pudiendo además incluir un carácter separador entre el texto de cada columna. Por ejemplo si tenemos una columna con el nombre y otra con los apellidos y queremos conseguir el formato 'apellido1 apellido2, nombre' lo podríamos hacer con esta opción. Es la acción opuesta a 'Split Column'.
- **Length**: permite obtener la longitud (número de caracteres) de la columna seleccionada.

- **Parse**: permite transformar una parte de texto en XML o JSON. Es una opción orientada a técnicos, no suele ser utilizada por usuarios.

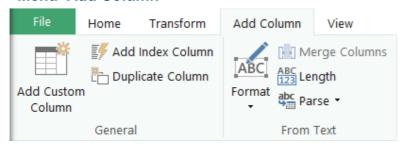
Grupo de opciones '*Number Column*': permiten realizar operaciones específicas sobre las columnas que contienen valores numéricos.

- **Statistics**: elimina las filas de la tabla y devuelve el resultado de la función utilizada (suma, mínimo, máximo, promedio, desviación standard, cuenta valores o cuenta valores distintos) para las filas de dicha tabla en base a la columna seleccionada. Su única utilidad es almacenar el dato resultante, sin ningún otro dato adicional.
- Standard: permite hacer una operación 'standard' (suma, resta, multiplicación, división, división entera y resto de la división) que afectará a todas las filas de la tabla, para la columna seleccionada. Por ejemplo nos valdría para incrementar el precio un 10%, para ello seleccionamos la columna 'precio', utilizamos la función standard 'Multiply' y como valor ponemos 1,10.
- Scientific: permite hacer una operación 'scientific' (valor absoluto, potencia, exponencial, raíz cuadrada, logaritmo, factorial) que afectará a todas las filas de la tabla, para la columna seleccionada.
- Trigonometry: permite hacer una operación trigonométrica (seno, coseno, tangente, arcoseno, arcocoseno, arcotangente) que afecta a todas las filas de la tabla, para la columna seleccionada.
- **Rounding**: permite hacer redondeos que afectarán a todas las filas de la tabla, para la columna seleccionada.
- Information: permite saber si el valor es par, impar, positivo, cero o negativo:
 - o **Is Even**: devuelve TRUE si el valor es par, FALSE en otro caso.
 - o **Is Odd**: devuelve TRUE si el valor es impar, FALSE en otro caso.
 - o **Sign**: devuelve 0 si el valor es 0, 1 si es positivo, -1 si es negativo.

Grupo de opciones '*From Date & Time*': permiten realizar operaciones específicas sobre las columnas que contienen fechas y horas.

- **Date**: permite aplicar diferentes funciones de fecha sobre una columna tipo '*Date*'. Por ejemplo: año, mes, día, último día del mes, etc.
- **Time**: permite aplicar diferentes funciones de hora sobre una columna tipo 'Date/Time' o 'Time'.
- **Duration**: permite calcular intervalos transcurridos, sólo aplica a columnas de tipo '*Duration*'.

Menu 'Add Column'



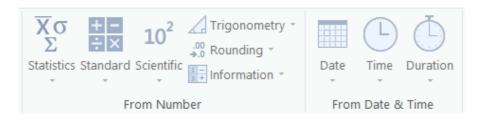


Figura 0-13 Ribbon, menú 'Add Column'.

Grupo de opciones '*General*': permiten realizar operaciones habituales a la hora de agregar nuevas columnas.

- Add Custom Column: permite crear columnas personalizadas, basadas en expresiones en Lenguaje M. Es una opción avanzada que no trataremos por el momento.
- Add Index Column: permite añadir una columna de índice, es decir, una columna, que empieza por el número cero y va asignando un número incremental a cada fila (es un autonumérico).
- **Duplicate Column**: hace una copia de la columna seleccionada manteniendo los mismos valores para sus filas.

Los grupos '*From Text*', '*From Number*' y '*From Date & Time*' ya han sido explicados anteriormente y se incluyen aquí también por usabilidad, evitando que tengamos que estar cambiando de menú con mayor frecuencia.

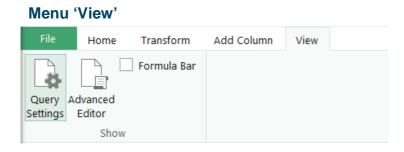


Figura 0-14 Ribbon, menú 'View'.

Grupo de opciones 'Show': permiten modificar ciertas configuraciones, mostrarlas y eliminarlas.

- **Query Settings**: muestra u oculta la ventana 'Query Settings' que aparece en la parte derecha de la ventana del 'Query Editor'.
- **Advanced Editor**: el editor avanzado muestra todos los pasos de la consulta actual en **Lenguaje M**.
- **Formula Bar**: muestra u oculta la barra de fórmulas (similar a la que tiene Excel para escribir las fórmulas).



Cada uno de los pasos que vamos definiendo para la obtención de datos y su posterior transformación, genera internamente una serie de expresiones en **Lenguaje M**, por tanto todo lo que vamos haciendo de forma visual aplicando transformaciones y demás, se podría hacer también escribiendo expresiones en Lenguaje M. Por ahora vamos a seguir trabajando sólo de forma visual, creando transformaciones con las opciones de menú que hemos estudiado, pero sí que dejaremos una imagen de qué se está generando por debajo para darle a conocer que ese código, si estudia dicho lenguaje lo podrá modificar directamente:

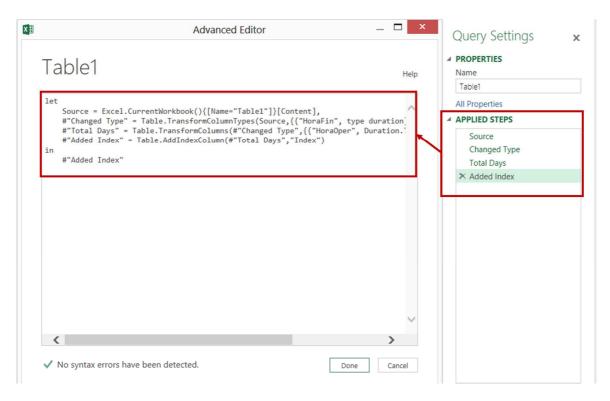


Figura 0-15 'Applied Steps' vs 'Advanced Editor'.

Muchas de estas opciones están también disponibles en los menús contextuales, y aparecerán o no en función del lugar donde hagamos clic y mostremos el menú. Aquí tiene una imagen de ejemplo:

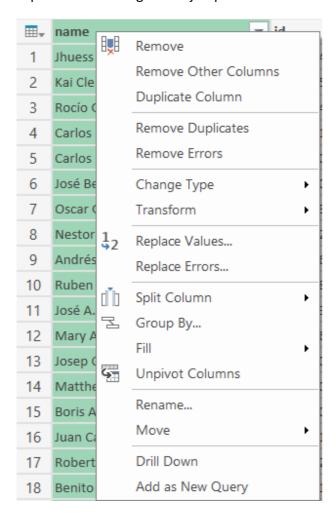


Figura 0-16 Menús contextuales.

Finalmente, veremos un ejemplo de uso en el que iremos realizando algunas transformaciones paso a paso.

- 1. Nos conectamos a SQL Server, a la base de datos 'TBI01 Origen'
- 2. Obtenemos los datos de la tabla 'Tiendas'
- 3. Ya en el 'Query Editor' eliminamos las columnas 'MovimientosTiendas', 'Presupuestos' y 'VentasCAB', mediante la opción '*Remove Columns*'.
- 4. Vemos que en la columna 'CodigosPostales' aparece un símbolo diferente CodigosPostales y en sus filas aparece la palabra 'Value', esto nos indica que es una tabla relacionada de la que podemos ir extrayendo columnas. Hacemos clic en h, seleccionamos las columnas 'CodioPostal' y 'Poblaciones' y pulsamos 'OK'.
- 5. Hacemos clic en , seleccionamos las columnas 'Poblacion' y 'Provincias' y pulsamos 'OK'.
- 6. Hacemos clic en http://, seleccionamos las columnas 'Provincia' y 'Paises' y pulsamos 'OK'.
- 7. Hacemos clic en , seleccionamos las columnas 'Poblacion' y 'Provincias' y pulsamos 'OK'.
- 8. Hacemos clic en , seleccionamos la columna 'Pais' y pulsamos 'OK'.
- 9. Eliminamos la columna 'CodigosPostales.CodioPostal'.
- 10. Renombramos la columna 'CodigosPostales.Poblaciones.Poblacion' como 'Poblacion'.
- 11. Renombramos la columna 'CodigosPostales.Poblaciones.Provincias.Provincia' como 'Provincia'.
- 12. Renombramos la columna 'CodigosPostales.Poblaciones.Provincias.Paises.Pais' como 'Pais'.

Hemos conseguido obtener en una sola tabla los datos de la tienda, junto con los datos geográficos (población, provincia, país) obtenidos de otras tablas relacionadas.

Por el momento lo dejamos aquí, con todas las transformaciones realizadas. Veamos una imagen en la que se muestra la tabla con el resultado obtenido, y en la parte derecha en la ventana 'Query Settings' la secuencia de pasos realizados 'Applied Steps', incluyendo la lectura desde el origen y todas las transformaciones realizadas.

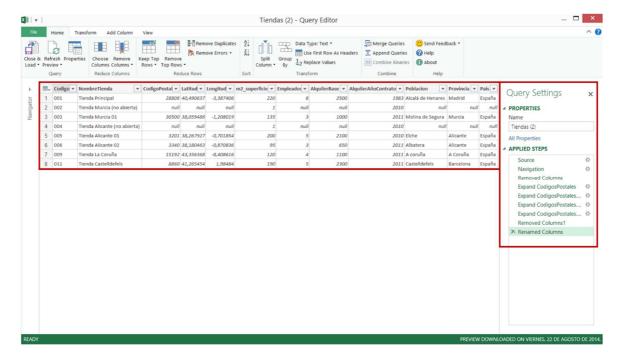


Figura 0-17 Resultado de la obtención de datos y las transformaciones realizadas.

La secuencia de pasos realizados queda disponible para ser modificada en cualquier momento, además de que será utilizada para ejecutarse cada vez que se pulse el botón 'Refresh', lo que producirá que se acceda de nuevo al origen de datos, se obtengan los datos que haya en ese momento y se apliquen todos los pasos con las transformaciones que hemos creado inicialmente.

Por último vamos a hacer una observación muy importante y que debe tener en cuenta. Cada vez que se ejecuta este conjunto de pasos, <u>al pulsar el botón 'Refresh' se elimina toda la información actual, dejando el destino vacío y volviéndolo a recargar.</u>

Destinos

Una vez que hemos conseguido las transformaciones necesarias para obtener el modelo analítico que necesitamos, tan sólo nos queda almacenar la información en el destino.

En este caso, Power Query tiene una gran limitación con respecto a las herramientas corporativas y de cara al uso por el departamento de TI: sólo puede almacenar el resultado de las transformaciones en una hoja Excel o en una tabla de Power Pivot del mismo libro en que hemos creado el proceso ETL.

Pero desde el punto de vista del BI Personal, que es el que estamos utilizando, realmente es justo lo que necesitamos. Estamos en un libro Excel, desde el que queremos automatizar una secuencia de pasos que nos permitan, cada vez que la ejecutemos, obtener la información que necesitamos para su análisis. Por tanto, Power Query es una herramienta que cubre perfectamente nuestras necesidades y esta limitación carece de importancia de cara al usuario final, que es a quien va dirigida.

Volviendo al ejemplo que hemos realizado anteriormente, vamos a continuar con los pasos finales:

13. Pulsamos el botón Load y vemos el resultado de una nueva tabla de tiendas que contiene los datos de las tiendas y los datos de su ubicación geográfica (Población, Provincia y Pais) que hemos obtenido de sus correspondientes tablas relacionadas.

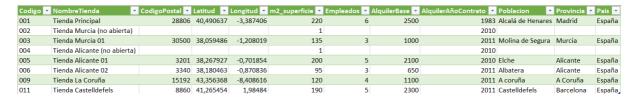


Figura 0-18 Tabla de la Dimension Tienda, incluye datos geográficos para el análisis.

También podemos pulsar en la parte inferior derecha de dicho botón y elegir entre una de estas dos opciones:

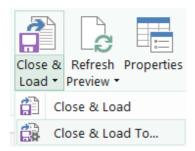


Figura 0-19 Opciones para guardar el resultado de las transformaciones en destino.

Si pulsamos 'Close & Load' guarda el resultado en una hoja Excel (Excel Worksheet).

Si pulsamos 'Close & Load To...' podemos elegir entre diversas opciones. Veamos la pantalla que nos aparece:

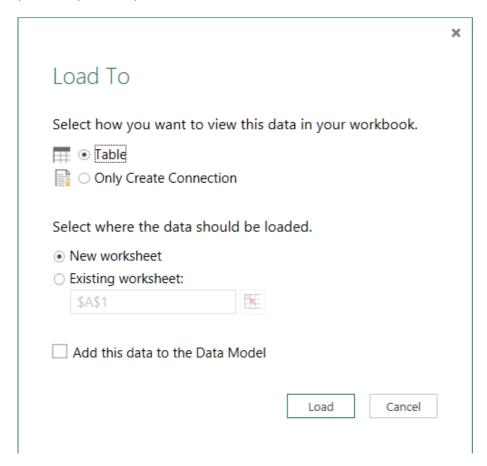


Figura 0-20 Load To...

- **Table**: guarda la información en una hoja Excel (*Excel Worksheet*) formateada como tabla.
- **Only Create Connection**: crea la conexión, pero no guarda nada en la hoja Excel.

En el caso de seleccionar anteriormente 'Table' puede elegir entre:

- **New Worksheet**: guardar en una nueva hoja.
- **Existing worksheet**: guardar en una hoja existente, especificando además la celda que quedará como esquina superior izquierda de la tabla.

En la parte inferior aparece un check:

 Add this data to the Data Model: esta es la opción que hace que los resultados se almacenen en Power Pivot (Excel Data Model). Este check es independiente del resto de opciones seleccionadas.

Teniendo en cuenta diversas combinaciones de las opciones anteriores, podemos almacenar la información sólo en una hoja Excel (bien en una existente o una nueva), sólo en Power Pivot, o combinar opciones y almacenar simultáneamente en Excel y Power Pivot.

Nuestra recomendación es que almacene los resultados sólo en Power Pivot y aproveche toda su potencia de cálculo y rapidez de respuesta.

Lenguaje M

Con todo lo estudiado hasta el momento puede considerarse un usuario intermedioavanzado de **Power Query**.

El **Lenguaje M** es un lenguaje orientado a la manipulación de datos, que le permite poder interactuar con Power Query y exprimir toda su potencia.

No es necesario utilizarlo en la mayoría de los casos. Pero, si más adelante desea profundizar más aún en la materia, debe estudiar dicho lenguaje y practicar hasta que domine la creación de expresiones con él, pudiéndolas utilizar tanto en el editor avanzado, como en la barra de fórmulas. Podrá crear consultas parametrizadas, y otra serie de alternativas que no son posibles utilizando solamente las opciones visuales (sin escribir código).

Power Query vs SSIS (SQL Server Integration Services)

Power Query es una herramienta *orientada al usuario*, mientras que **SSIS** es una herramienta **orientada a desarrolladores y personal de TI**.

SSIS tiene una mayor funcionalidad, mayor cantidad de transformaciones y destinos, mayor flexibilidad. Es mucho más escalable, permitiendo un mayor volumen de datos y un aumento periódico de estos mayor. Tiene capacidades de programación en VB.NET y C#. Es una herramienta extensible, el desarrollador puede crear nuevos orígenes, transformaciones y destinos e integrarlos en la herramienta.

Power Query tiene ciertas limitaciones que debe tener en cuenta:

- No permite carga periódica automatizada. La forma de volver a cargar la información es pulsando manualmente el botón "Refresh".
- <u>No permite hacer cargas incrementales</u>, manteniendo información que ya había sido incorporada, el proceso de refresco (*Refresh*) hace un borrado total y una recarga completa.
- Aunque tiene una amplia variedad de transformaciones e incluso un lenguaje propio, Lenguaje M, la funcionalidad es más limitada que la de SSIS.
- Como destino sólo podemos tener archivos .xlsx, almacenando bien en hojas Excel tradicionales, bien en Power Pivot.

Si tenemos una herramienta como SSIS, ¿qué sentido tiene crear Power Query?

Básicamente quitar dependencia del usuario del departamento de TI y permitirle obtener datos de forma independiente. Es muy útil tanto para el BI Personal como para el Self-Service BI. Son herramientas muy diferentes orientadas a públicos diferentes. Son herramientas que se complementan, Power Query viene a cubrir un

hueco muy demandado por los usuarios de negocio desde hace tiempo, que les permite obtener datos y convertirlos en información por ellos mismos.

Conclusiones

Como ha podido comprobar **Power Query**, aunque con algunas limitaciones comparado con herramientas ETL corporativas, tiene una inmensa cantidad de posibilidades, tanto a la hora de acceder a muy diversos orígenes de datos, como para aplicar transformaciones sobre ellos y convertirlos en un modelo que nos facilite su análisis, aplicando las buenas prácticas estudiadas en capítulos anteriores.

Posiblemente se haya quedado con la sensación de que hay una gran cantidad de posibilidades, que las entiende por separado, pero que una cosa muy diferente es ponerlas en práctica, y cuando tenemos un caso real, ser capaz de estudiar el origen, extraer de él los datos que necesitamos, hacer un buen número de transformaciones y cálculos, hasta conseguir un modelo que nos permita realizar un buen análisis sobre ellos y finalmente guardarlos en Power Pivot, para posteriormente mejorar y enriquecer dicho modelo.

Si está realizando alguno de los <u>cursos o Masters de SolidQ</u>, esa sensación habrá desaparecido, o al menos disminuido enormemente ©, ya que en ellos se realizan una serie de "laboratorios guiados paso a paso" para consolidar lo estudiado sobre diversos casos prácticos reales.

Si no es su caso, le recomiendo que se forme con nuestro curso:

Analiza tu Negocio con Excel y Power Bl. Aprende de tus datos

Quiero obtener información detallada y registrarme

ENLACES ESENCIALES

Visita el <u>blog de Salvador Ramos</u>

Mi libro "Microsoft BI: Vea el cubo medio lleno" (descárgalo gratis)

Linkedin Salvador Ramos

Twitter Salvador Ramos

Visita los blogs de SolidQ

Libros y publicaciones de SolidQ, la mayoría gratuitos (descárgalos aquí)

Accede a los Cursos de SolidQ

Accede al Calendario de cursos de SolidQ (continuamente actualizado)

Accede a los Servicios de SolidQ

Twitter de SolidQ, y también nuestro canal en español

GRACIAS

Antes de que te vayas, quiero decirte "Gracias por leer mi libro".

Sé que hay otros muchos libros en la materia, pero apostaste por este. Ahora sólo quiero pedirte un "pequeño" favor:

Dedica unos minutos y deja un testimonio en Amazon, lo necesito para seguir mejorando y aportando contenido que sea de tu interés.

Y, por último, si crees que merece la pena compartir este libro, ¿podrías tomarte unos segundos y mostrárselo a tus seguidores en las Redes Sociales? El boca a boca es crucial para hacer llegar estos conocimientos al mayor número posible de personas interesadas en la materia, les serás de gran ayuda. Si tienes un momento, te estaré muy agradecido.

¡Gracias!