



Caso 140 Conexión a carpeta “Facturas”

01 ETL y Modelo de Datos. PQ Doble conexión a carpeta con transformación. Crear consulta combinada

Jose Ignacio González Gómez

Departamento de Economía, Contabilidad y Finanzas - Universidad de La Laguna

www.jggomez.eu

R25

Ejercicio adaptado de: [Taller - Limpieza de datos con Power BI](#)

Contenido

Resumen y objetivos	2
1 Aspectos técnicos relacionados con la conexión a una carpeta	3
1.1 Objetivo y ejemplo de uso	3
1.2 Consideraciones previas.....	4
1.3 Las conexiones a las carpetas, funcionamiento básico.....	5
1.3.1 Selección de la carpeta y del fichero de ejemplo o de referencia	5
1.3.2 Carpeta “Transformar archivo de..”	6
1.3.3 Subcarpeta Consulta Auxiliar (A).....	6
1.3.4 Otras Consultas (B).....	7
1.4 Conveniencia de crear un parámetro de conexión	7
2 Presentación del caso “Facturas”	8
2.1 Introducción, conexiones a una carpeta	8
2.2 El modelo de factura y su contenido.....	9
2.3 Se pide	10
3 Primeros pasos en el proceso ETL.....	11
3.1 Creamos la conexión a la carpeta y combinamos y transformamos (Facturas Consolidadas v1 Conexion y parametro)	11
3.2 Creando una segunda conexión Factura Detalle	14
3.3 Creando el parámetro de conexión y aplicando	15
4 Transformaciones básicas	17
4.1 Transformación de la consulta “Cabecera” (Facturas Consolidadas v2a ETL Cabecera)	17
4.1.1 Transformar archivo de ejemplo, objetivo.....	17
4.1.2 Limpieza de 1º Grado-Técnicas de depuración.....	17
4.1.3 Resumen del proceso Limpieza de Datos de 1º Grado, Técnicas de Depuración	19
4.2 Transformación de la consulta “Detalle” (Facturas Consolidadas v2b ETL Detalle) ...	19

4.2.1	Transformar archivo de ejemplo (2), objetivo	19
4.2.2	Limpieza de 1º grado (depuración) y Limpieza de 2º grado (Dinamización por transposición).....	19
4.2.3	Resumen del proceso Limpieza de 1º grado y 2º grado	22
5	Consulta combinada Cabecera y Detalle (Facturas Consolidadas v3 Combina Cabecera_Detalle).....	23
6	Resultado final objetivo	25

Resumen y objetivos

Power Query. Creando parámetros de conexión

Conexión a carpeta: doble conexión para dos procesos ETL distinto para cabecera de factura y detalle. Crear consulta combinada

Proceso Transformación: Aplicación de técnicas de depuración sobre filas y columnas basicas como formateo, encabezado promovidos, tipos cambiado, quitar filas en blanco, filtros (quitar totales, null...), quitar columnas vacías, dividir columnas, etc.

Técnicas avanzadas: transposición, combinar consultas, etc...

1 Aspectos técnicos relacionados con la conexión a una carpeta

1.1 Objetivo y ejemplo de uso

La conexión a una carpeta tiene como objetivo principal importar y combinar archivos que se encuentran en una misma carpeta. Esto es especialmente útil cuando disponemos de múltiples archivos con la misma estructura (como informes mensuales) y deseamos consolidar todos esos datos en una sola tabla para su análisis. Esto te permite:

1. **Importar varios archivos al mismo tiempo:** Si tenemos una carpeta con varios archivos del mismo tipo (como Excel, CSV, etc.), Power Query puede combinarlos en un solo conjunto de datos.
2. **Automatizar la actualización de datos:** Si los archivos en la carpeta se actualizan o se agregan nuevos archivos, el editor de consulta puede procesar y actualizar los datos automáticamente sin necesidad de ajustar cada archivo individualmente.
3. **Ahorro de tiempo:** Evita la labor tediosa de importar y transformar cada archivo por separado, lo que es especialmente útil para grandes volúmenes de datos.
4. **Consistencia en el procesamiento de datos:** Todos los archivos en la carpeta se procesan de la misma manera, asegurando que los datos sigan el mismo formato y estructura.

En resumen, la conexión a una carpeta en Power Query es una herramienta poderosa para automatizar la importación, combinación y transformación de datos de múltiples archivos, lo que facilita el análisis y la toma de decisiones informadas.

Presentamos a continuación algunos ejemplo de casos de uso:

Análisis de ventas

Nos permite combinar datos de ventas de diferentes archivos mensuales en un solo informe. Por ejemplo, una empresa genera un informe de ventas en formato Excel cada mes. Conectar Power Query a una carpeta donde se almacenan estos archivos permite combinar todos los informes mensuales en un solo informe acumulado

Consolidación de datos de encuestas

Unir respuestas de encuestas almacenadas en múltiples archivos CSV, es decir combinar todas las respuestas en un único conjunto de datos, ahorrando tiempo y esfuerzo.

Finanzas y Contabilidad

Para empresas que reciben extractos bancarios, reportes financieros, balances por sucursales, archivos de texto con transacciones, bases de datos con información de clientes, etc.. en múltiples archivos Excel mensuales, una conexión a carpeta permite reunir toda la información financiera en un único lugar para análisis y reconciliación.

Integración con Datos de Proveedores

Por ejemplo, varios proveedores envían archivos de precios de productos y estos se almacenan en una carpeta. Power Query permite combinarlos y actualizar automáticamente la base de datos de precios.

Datos de Sensores

Si estamos trabajando con datos de sensores que generan archivos CSV cada hora o cada día, podemos usar una conexión a carpeta para consolidar todos estos datos en una sola tabla, facilitando el análisis.

Investigación y Análisis de Datos:

En proyectos de investigación donde se recolectan datos en archivos individuales (por ejemplo, estudios de campo con múltiples registros), conectar una carpeta en Power Query facilita la integración y análisis de todos los datos recogidos.

Seguimiento de Inventarios de Varias Sucursales

Empresas que manejan inventarios en múltiples ubicaciones pueden tener archivos de inventario separados por ubicación. Al conectar una carpeta en Power Query, podemos consolidar todos estos archivos en un solo inventario maestro.

1.2 Consideraciones previas

Antes de empezar recomendamos tomar en consideración los siguientes aspectos:

- Asegúrese de que todos los archivos que desea combinar están contenidos en una carpeta dedicada sin archivos extraños. En caso contrario, todos los archivos de la carpeta y las subcarpetas que seleccione se incluirán en los datos que se combinarán.
- Cada archivo debe tener el mismo esquema con encabezados de columna coherentes, tipos de datos y número de columnas. Las columnas no tienen que estar en el mismo orden que la coincidencia se realiza mediante nombres de columna.
- Si es posible, evite objetos de datos no relacionados para orígenes de datos que pueden tener más de un objeto de datos, como un archivo JSON, un libro de Excel o una base de datos de Access

Todos los pasos de limpieza y transformación que aplicaremos al fichero ejemplo o de referencia se reproducirán al resto de ficheros que añadiremos.

Procedemos por tanto a realizar los ajustes de transformación aplicando las técnicas de limpieza de primer grado y de segundo grado más frecuentes como la del esquema:

<p>Limpieza de Datos de Primer Grado: <i>Técnicas de Depuración</i></p>	<ol style="list-style-type: none"> 1. Reducción <ul style="list-style-type: none"> ▪ Exclusión de filas ▪ Quitar columnas ▪ Promover encabezado 2. Formateo de Datos <ul style="list-style-type: none"> ▪ Transformación a nivel de columnas ▪ Reemplazar valores ▪ Dividir columnas ▪ Asignación de nombres 3. Transposición 4. Rellenar
<p>Limpieza de Datos de Segundo Grado: <i>Técnicas de Dinamización por Transposición</i></p>	<ol style="list-style-type: none"> 1. Anulación de Dinamización (Unpivot) 2. Columna Dinámica (Pivot)

1.3 Las conexiones a las carpetas, funcionamiento básico.

1.3.1 Selección de la carpeta y del fichero de ejemplo o de referencia

Expondremos a continuación cómo funciona las conexiones a las carpetas. Cuando se realiza una conexión a una carpeta supone que deseamos combinar o anexar el contenido de los distintos ficheros que contiene la misma, lo cual exige que estos no solo sean del mismo tipo (pdf, xls, csv...) sino que además tengan la misma estructura o composición, como son el número y tipo de columnas, el número de tablas, etc... En el caso de que existan ficheros de distinto tipo tendremos que seleccionar un único tipo de fichero (pdf, xls, csv...) con el que queremos trabajar.

Una vez creada la conexión a, tal y como muestra la Ilustración 1 se nos presenta la relación de ficheros disponibles, proponiendo como fichero ejemplo o de referencia el primero, pero podríamos seleccionar otro de los disponibles. Este será el fichero de ejemplo sobre el que se definirán las transformaciones necesarias que se replicarán al resto de ficheros.

Señalar que sería conveniente seleccionar aquel fichero que requiera de más procesos y técnicas de depuración y transformación, por tanto, sería conveniente llevar a cabo un análisis de los ficheros disponibles y de su contenido.

Una vez seleccionado el fichero de ejemplo, el editor de consulta analiza su contenido e identifica los distintos elementos que lo componen (Parámetro), en nuestro caso a identificado 3 elementos, 2 tablas y la página completa.

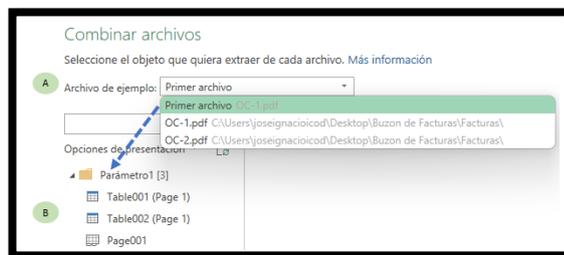


Ilustración 1

Podremos visualizar la información disponible en cada elemento para seleccionar según su contenido, el que contiene la información que nos interesa, si esta se encuentra en una tabla concreta o en toda la página.

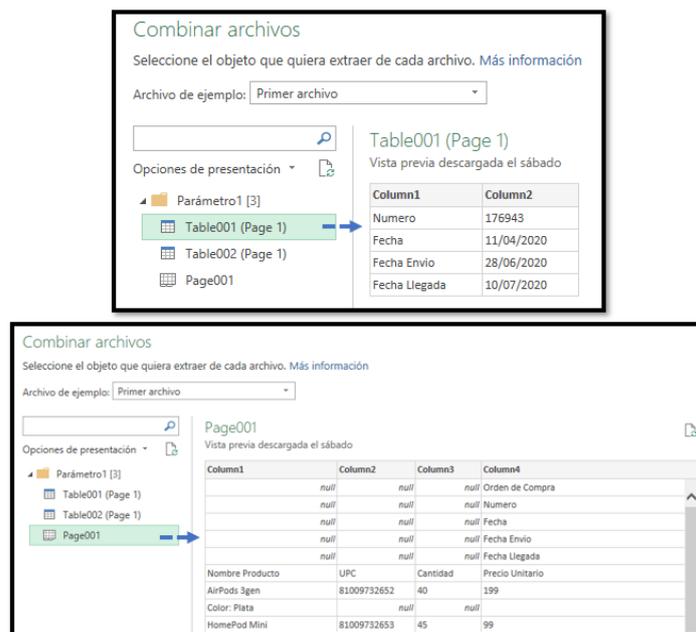


Ilustración 2

El resultado de la conexión es la creación de dos estructuras “Transformar archivo de...” y “Otras Consultas”, ambas vinculadas y que pasamos a analizar (ver Ilustración 3).

1.3.2 Carpeta “Transformar archivo de..”

Una vez seleccionado el archivo de ejemplo y el elemento (tabla o página) que contiene la información de interés, el editor de consultas genera una estructura para cada conexión que la agrupa en la carpeta “Transformar archivo de Facturas [2]” donde el nombre se corresponde con el nombre de la carpeta a la que estamos conectando (Facturas, por ejemplo) y lo añade a la raíz “Transformar archivo de” generando así la carpeta de conexión.

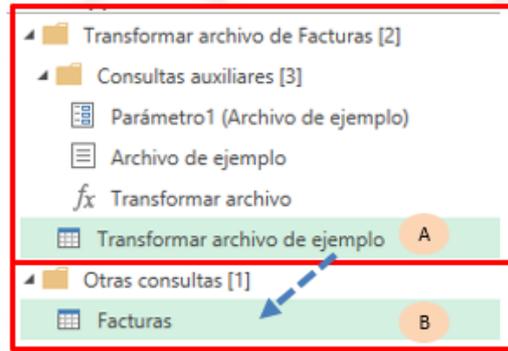


Ilustración 3

El numero 2 se refiere al número de elementos disponibles en la parametrización de la conexión y que son:

1. La subcarpeta Consultas Auxiliares [3], que contiene tres elementos relacionados con los parámetros de configuración que definen la consulta.
2. El fichero “Transformar archivo de ejemplo” (A), es una copia del seleccionado en la Ilustración 2 y sobre el que se aplicarán todos los ajustes de limpieza y transformación generando una plantilla que se replicarán uno a uno al resto de ficheros contenidos en la carpeta de conexión.

1.3.3 Subcarpeta Consulta Auxiliar (A)

Centrándonos en la subcarpeta Consultas Auxiliares esta contiene tres elementos (A, B y C):

- A. Parámetro 1 en el que se define un parámetro que toma como valor “Archivo de Ejemplo” podríamos cambiarlo como, por ejemplo, fichero de referencia.
- B. Identifica al fichero seleccionado en la Ilustración 1.
- C. Es una función interna creada.
- D. Como señalamos el fichero “Transformar archivo de ejemplo” contiene la plantilla de ajustes de limpieza y transformación que se replicaran a cada fichero contenidos en la carpeta de conexión.

Es importante señalar que si creamos un parámetro de conexión debemos actualizarlo en los siguientes elementos en el apartado origen del archivo: B.

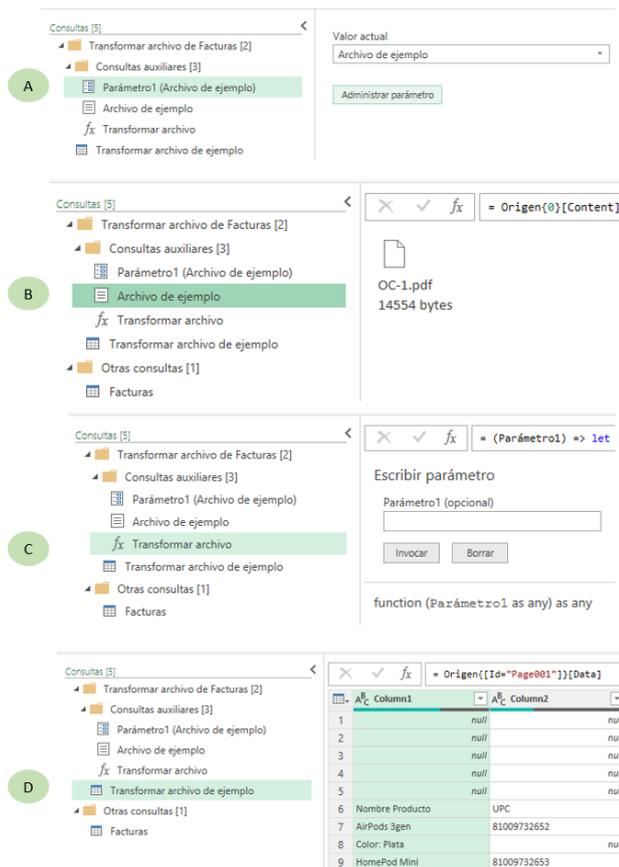


Ilustración 4

1.3.4 Otras Consultas (B)

En el momento de crear la conexión tal y como hemos mostrado en la Ilustración 3 se ha creado además otra carpeta con el nombre “Otras consultas”, la cual contiene una consulta de agregación consolidación de la información contenido en el resto de ficheros disponibles en la carpeta a la que estamos conectado, y a cada uno de estos ficheros se les ha aplicado la plantilla de limpieza y depuración, s.

	Source.Name	Column1	Column2
1	OC-1.pdf	null	null
2	OC-1.pdf	null	null
3	OC-1.pdf	null	null
4	OC-1.pdf	null	null
5	OC-1.pdf	null	null
6	OC-1.pdf	Nombre Producto	UPC
7	OC-1.pdf	AirPods 3gen	81009732652
8	OC-1.pdf	Color: Plata	null
9	OC-1.pdf	HomePod Mini	81009732653
10	OC-1.pdf	Color: Azul	null
11	OC-1.pdf	TotalMount Pro Max	81009732654
12	OC-1.pdf	Color: Negro	null

Ilustración 5

Así y según la Ilustración 3, **A** contiene la plantilla con todos los ajustes necesarios de limpieza y transformación y la consulta **B** es el resultado de aplicar esta plantilla a todos los ficheros contenidos en la carpeta de conexión, y según el caso se han consolidado en una sola tabla.

En resumen

De esta forma los ajustes de limpieza y transformación que se definan sobre el fichero de ejemplo o de referencia serán replicados al resto para proceder posteriormente a combinar o anexar el contenido de cada uno de ellos en uno solo que será el resultado u objetivo de la conexión a la carpeta.

1.4 Conveniencia de crear un parámetro de conexión

Es recomendable contar con un parámetro de conexión y aplicarlo a las distintas conexiones de carpetas que hayamos creado ya que esto permite mejorar significativamente la flexibilidad en la actualización del origen de datos al cambiar fácilmente las fuentes (como rutas de archivos o nombres de servidores) sin tener que modificar cada consulta individualmente lo que hace que el mantenimiento sea más sencillo.

En la siguiente ilustración hemos creado el parámetro de conexión denominado “ConexiónCarpeta” y que contiene la ruta a la fuente de datos (carpeta que nos interesa), así una vez creado el parametro tendremos que cambiar el origen de las 2 consultas al citado parámetro, Archivo de ejemplo y Facturas, tal y como se muestra en la Ilustración 6

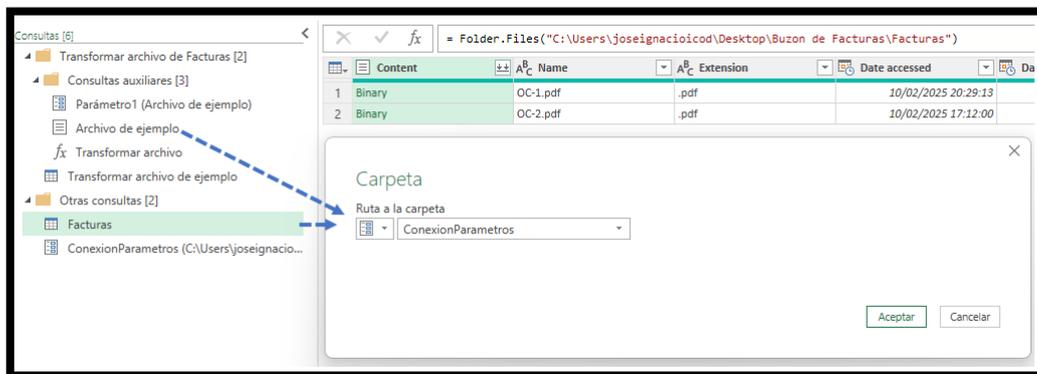


Ilustración 6

2 Presentación del caso “Facturas”

Fuente: [Taller - Limpieza de datos con Power BI](#)

2.1 Introducción, conexiones a una carpeta

Disponemos de una carpeta “Facturas” que contiene dos ficheros pdf que se corresponde con facturas de un mismo proveedor y por tanto tienen el mismo formato o estructura. También disponemos de otra carpeta Facturas Pdtes que contiene otras facturas del mismo proveedor. Queremos realizar un proceso ETL de conexión a la carpeta inicial Facturas y realizar las transformaciones necesarias para convertir los datos contenidos en los ficheros no normalizado en uno correctamente tabulado.

A Cabecera				
Orden de Compra	Numero			176943
	Fecha			11/04/2020
	Fecha Envio			28/06/2020
	Fecha Llegada			10/07/2020

B Detalle				
Nombre Producto	UPC	Cantidad	Precio Unitario	Total Linea
AirPods 3gen	81009732652	40	199	7,960.00
Color: Plata				
HomePod Mini	81009732653	45	99	4,455.00
Color: Azul				
TotalMount Pro Max	81009732654	40	35.95	1,438.00
Color: Negro				
Soporte Magnético Belkin	81009732655	50	34.95	1,747.50
Color: Negro				
Funda Smart Folio	81009732656	100	89	8,900.00
Color: Azul Mar				
Etiqueta para maleta	81009732658	45	449	20,205.00
Color: Marrón				
			Subtotal	44,705.50
			Total	44,705.50

C Pie
Notas
Por favor confirmar recepción al correo. Gracias

Ilustración 7

Este proceso ETL exigirá realizar diversos ajustes aplicando técnicas de depuración sobre filas y columnas básicas como formateo, encabezado promovidos, tipos cambiado, quitar filas en blanco, filtros (quitar totales, null...), quitar columnas vacías, dividir columnas, rellenar, etc así como técnicas avanzadas de trasposición...

<p>Limpieza de Datos de Primer Grado: <i>Técnicas de Depuración</i></p>	<ol style="list-style-type: none"> Reducción <ul style="list-style-type: none"> Exclusión de filas Quitar columnas Promover encabezado Formateo de Datos <ul style="list-style-type: none"> Transformación a nivel de columnas Reemplazar valores Dividir columnas Asignación de nombres Trasposición Rellenar
<p>Limpieza de Datos de Segundo Grado: <i>Técnicas de Dinamización por Trasposición</i></p>	<ol style="list-style-type: none"> Anulación de Dinamización (Unpivot) Columna Dinámica (Pivot)

Esquema 1

2.2 El modelo de factura y su contenido

Como hemos señalado, los documentos corresponden a las facturas de un mismo proveedor y tienen una estructura basada en tres apartados:

- A. La **cabecera** que contiene todos los campos básicos relacionadas con la factura.
- B. **Detalle**, donde se dispone de los productos/servicios facturados, destacando que existe una segunda línea en la que se identifica el color del producto.
- C. **Pie** de factura en la que se recoge las notas relacionadas con el documento.

A Cabecera					
Orden de Compra		Numero			176943
		Fecha			11/04/2020
		Fecha Envio			28/06/2020
		Fecha Llegada			10/07/2020

B Detalle				
Nombre Producto	UPC	Cantidad	Precio Unitario	Total Linea
AirPods 3gen	81009732652	40	199	7,960.00
Color: Plata				
HomePod Mini	81009732653	45	99	4,455.00
Color: Azul				
TotalMount Pro Max	81009732654	40	35.95	1,438.00
Color: Negro				
Soporte Magnético Belkin	81009732655	50	34.95	1,747.50
Color: Negro				
Funda Smart Folio	81009732656	100	89	8,900.00
Color: Azul Mar				
Etiqueta para maleta	81009732658	45	449	20,205.00
Color: Marrón				
Subtotal				44,705.50
Total				44,705.50

C Pie
Notas
Por favor confirmar recepción al correo. Gracias

Ilustración 8

La información que interesa es la de la cabecera y detalle, por tanto, este es un caso peculiar donde debemos realizar dos transformaciones por separado, una para la cabecera y otro para el detalle y por tanto va a exigir contar con dos conexiones que generaran dos carpetas de "Transformación de archivo de ..." con su consulta auxiliar respectiva (una para cada objetivo) transformación lo que implica por tanto el crear dos consultas a la carpeta que va a contener los ficheros.

Obvia decir que todos los ficheros deben tener la misma estructura de datos, cuestión que se cumple en nuestro caso del analisis de los mismos y que el nombre del fichero debe ser unico.

2.3 Se pide

Una vez finalizado todo el proceso ETL de conexión a la carpeta Facturas disponer en nuestro documento de trabajo (Excel o Power BI) de tres pestañas correspondientes a las consultas creadas, tal y como se muestra en la siguiente Ilustración.

The image displays three screenshots of an Excel spreadsheet, illustrating different levels of data detail for invoices.

Top Screenshot: Facturas Cabecera

	A	B	C	D	E	F
1	Source.Name	Numer	Fecha	Fecha Envio	Fecha Llegada	
2	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020	
3	OC-2.pdf	159753	11/04/2020	28/06/2020	10/07/2020	
4						

Middle Screenshot: Facturas Detalles

	A	B	C	D	E	F	G
1	Source.N	Producto	Color	Codigo	Cantid	Precio Unitar	Total Linea
2	OC-1.pdf	AirPods 3gen	Plata	81009732652	40	199,00	7960,00
3	OC-1.pdf	HomePod Mini	Azul	81009732653	45	99,00	4455,00
4	OC-1.pdf	TotalMount Pro Max	Negro	81009732654	40	35,95	1438,00
5	OC-1.pdf	Soporte Magnético Belkin	Negro	81009732655	50	34,95	1747,50
6	OC-1.pdf	Funda Smart Folio	Azul Mar	81009732656	100	89,00	8900,00
7	OC-1.pdf	Etiqueta para maleta	Marrón	81009732658	45	449,00	20205,00
8	OC-2.pdf	Mochila de cuero	Marrón	81893502692	50	35,80	1790,00
9							

Bottom Screenshot: Cabecera y Detalle

	A	B	C	D	E	F	G	H	I	J	K
1	Source.Name	Producto	Color	Codigo	Cantidad	Precio Unitario	Total Linea	Numero	Fecha	Fecha Envio	Fecha Llegada
2	OC-1.pdf	AirPods 3gen	Plata	81009732652	40	199	7960	176943	11/04/2020	28/06/2020	10/07/2020
3	OC-1.pdf	HomePod Mini	Azul	81009732653	45	99	4455	176943	11/04/2020	28/06/2020	10/07/2020
4	OC-1.pdf	TotalMount Pro Max	Negro	81009732654	40	35,95	1438	176943	11/04/2020	28/06/2020	10/07/2020
5	OC-1.pdf	Soporte Magnético Belkin	Negro	81009732655	50	34,95	1747,5	176943	11/04/2020	28/06/2020	10/07/2020
6	OC-1.pdf	Funda Smart Folio	Azul Mar	81009732656	100	89	8900	176943	11/04/2020	28/06/2020	10/07/2020
7	OC-1.pdf	Etiqueta para maleta	Marrón	81009732658	45	449	20205	176943	11/04/2020	28/06/2020	10/07/2020
8	OC-2.pdf	Mochila de cuero	Marrón	81893502692	50	35,8	1790	159753	11/04/2020	28/06/2020	10/07/2020
9											
10											

Ilustración 9

3 Primeros pasos en el proceso ETL

3.1 Creamos la conexión a la carpeta y combinamos y transformamos (Facturas Consolidadas v1 Conexion y parametro)

Fichero: Facturas Consolidadas v1 Conexion y parametro

Comenzamos creando nuestro fichero de trabajo (en Excel o Power BI) y accedemos al editor de consultas (Power Query) para crear nuestra primera conexión a la fuente de datos, en este caso es a la carpeta que contiene las facturas (Ilustración 10).

Se nos presenta un nuevo cuadro de dialogo (Ilustración 11) que detecta los dos ficheros que contiene la carpeta Facturas, solicitando la acción que queremos ejecutar si queremos *combinar y transformar datos* o solo *transformar*.

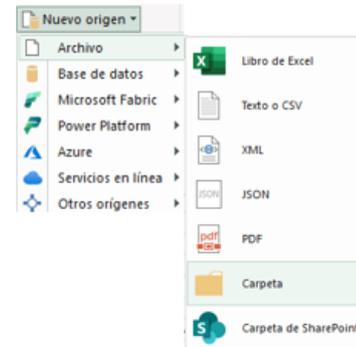


Ilustración 10

Como nuestro interes es consolidar la información disponibles de todos los ficheros elegimos "*Combinar y transformar datos*" que nos lleva a un nuevo cuadro de dialogo "Combinar archivos", en el que en primer lugar debemos seleccionar que fichero tomaremos como **Archivo de ejemplo**, si el primero de la lista o alguno en concreto de los disponibles en la carpeta (ver Ilustración 11).

En este punto es importante tomar en consideración lo que ya hemos manifestado al respecto en el sentido que *sería conveniente seleccionar aquel fichero que requiera de más procesos y técnicas de depuración y transformación*.

Esto habilita la opcion de "Combinar" y del analisis de las tablas, podemos observar que toda la información que nos interesa está en "Page001" y es con esta con la que vamos a trabajar para llevar a cabo el proceso de transformación

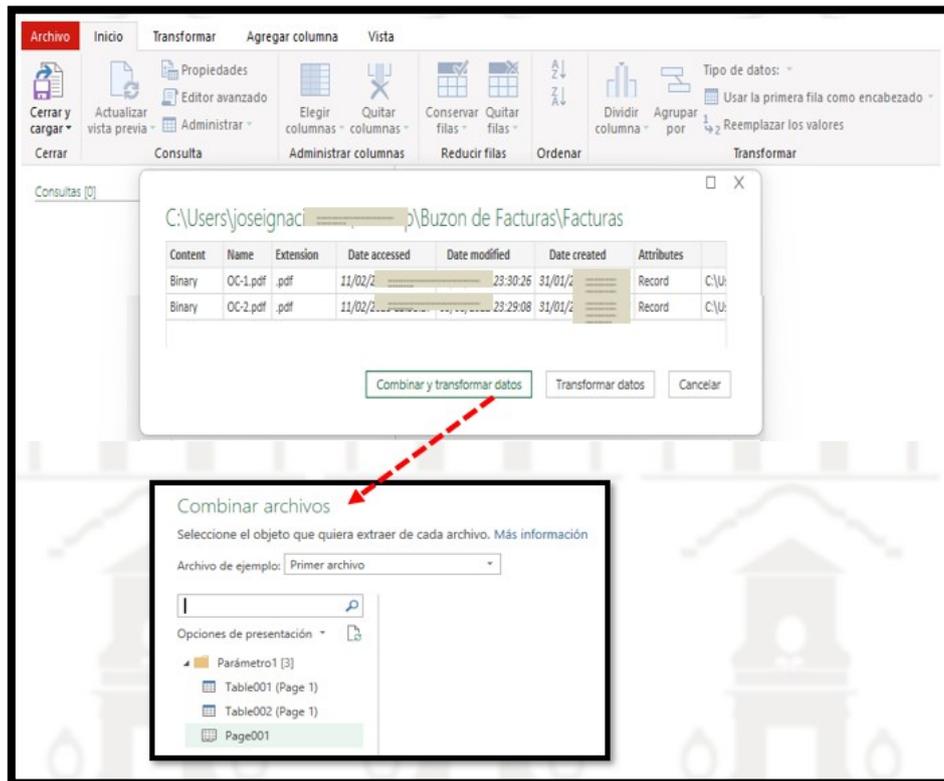


Ilustración 11

Una vez seleccionado el *Archivo de ejemplo*, analiza su contenido identificando las tablas de y páginas con información detectadas con el fin de que podamos previsualizar y seleccionar el objeto (tabla o página) sobre el que vamos a realizar el proceso de depuración y transformación y se generara la plantilla de referencia con los ajustes requeridos.

Observamos en la Ilustración 12, la previsualización del contenido de cada uno de los objetos, en concreto, la Table001 detectada en la página 1 del documento contiene los valores mostrados a la derecha que se corresponde con el encabezado de la factura. Igualmente, a detectado una segunda tabla, Table002 en la página 1 y según la previsualización se corresponde con el detalle de la factura y finalmente nos muestra la posibilidad de seleccionar el contenido total de la pagina1 (Pag001)

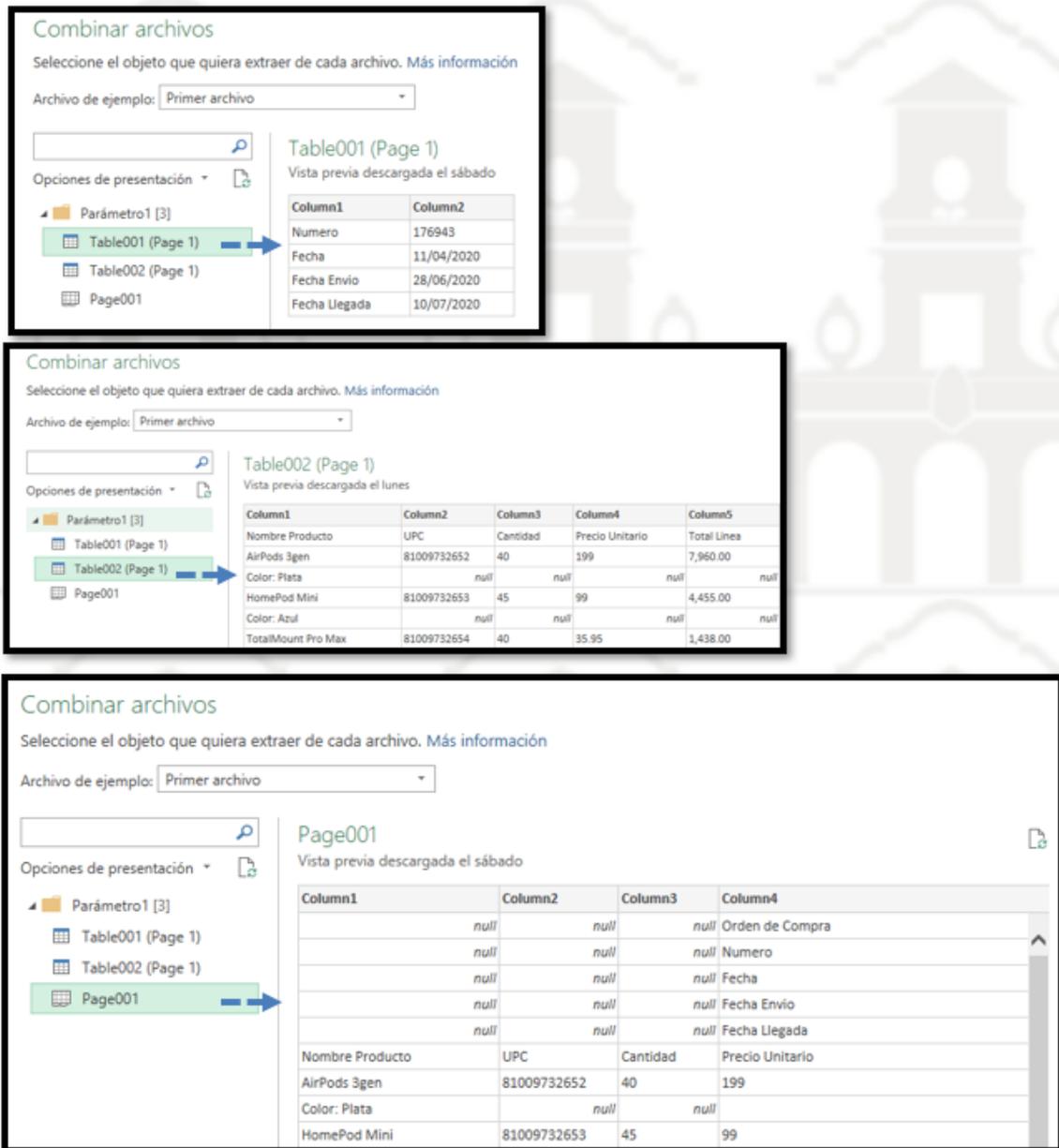
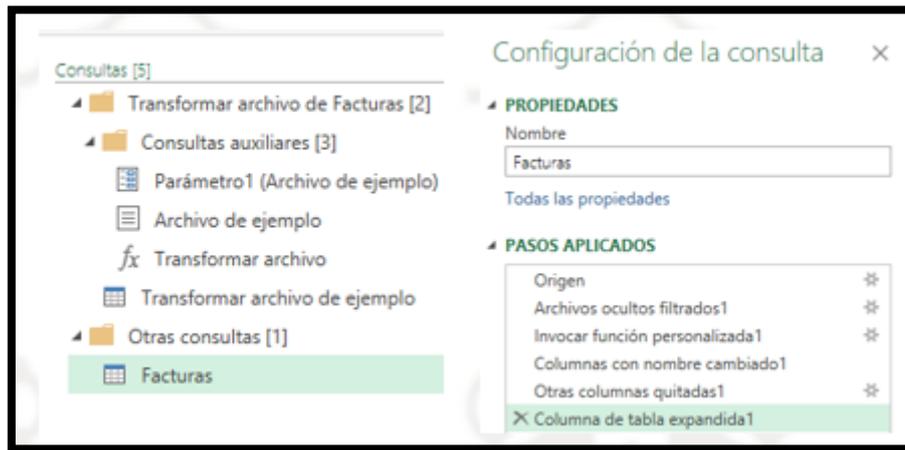


Ilustración 12

En nuestro caso nos hemos decantado por trabajar sobre el objeto Page001, pero podríamos seleccionar la tabla que contiene la cabecera de las facturas o la tabla con el detalle. En este caso disponemos de toda la información.

Este es el resultado de esta primera parte de la conexión donde se ha creado, como hemos comentado, una estructura de carpetas así como los pasos aplicados de configuración obteniendo como resultado una consulta donde disponemos de toda la información necesaria y sobre la cual vamos a aplicar las transformaciones básicas.



	Source.Name	Column1	Column2	Column3	Column4	Column5
1	OC-7.pdf		null	null	null	Orden de Compra
2	OC-7.pdf		null	null	null	Numero
3	OC-7.pdf		null	null	A	Fecha
4	OC-7.pdf		null	null	null	Fecha Envío
5	OC-7.pdf		null	null	null	Fecha Llegada
6	OC-7.pdf	Nombre Producto	UPC	Cantidad	Precio Unitario	Total Linea
7	OC-7.pdf	AirPods Sgen	81009732652	40	199	7,960.00
8	OC-7.pdf	Color: Plata		null	null	null
9	OC-7.pdf	HomePod Mini	81009732653	45	99	4,455.00
10	OC-7.pdf	B	Color: Azul	null	null	null
11	OC-7.pdf	TotalMount Pro Max	81009732654	40	35.95	1,438.00
12	OC-7.pdf	Color: Negro		null	null	null
13	OC-7.pdf	SopORTE Magnético Belkin	81009732655	50	34.95	1,747.50
14	OC-7.pdf	Color: Negro		null	null	null
15	OC-7.pdf	Funda Smart Folio	81009732656	100	89	8,900.00
16	OC-7.pdf	Color: Azul Mar		null	null	null
17	OC-7.pdf	Etiqueta para maleta	81009732658	45	449	20,205.00
18	OC-7.pdf	Color: Marrón		null	null	null
19	OC-7.pdf		null	null	null	Subtotal
20	OC-7.pdf		null	null	null	Total
21	OC-7.pdf	C	Notas	null	null	null
22	OC-7.pdf		null	null	null	Por favor confirmar recepción al correo. Gracias

Ilustración 13

Este es el resultado de esta primera parte de la conexión donde se ha creado, como hemos comentado, una estructura de carpetas así como las pasos aplicados de configuración obteniendo como resultado una consulta donde disponemos de toda la información necesaria y sobre la cual vamos a aplicar las transformaciones básicas. Para finalizar a esta consulta le podemos cambiar el nombre como, por ejemplo, Cabecera

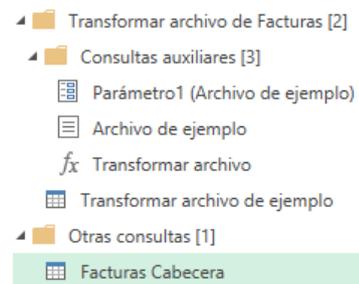


Ilustración 14

3.2 Creando una segunda conexión Factura Detalle

Como hemos comentado nos interesa realizar dos transformaciones una para la cabecera y otra para el detalle, pero no podemos actuar sobre el mismo objeto (Page001) porque algunas de las transformaciones aplicadas para obtener la cabecera afectarían al contenido del detalle al haber eliminado filas y columnas de contenido relacionado y por tanto no sería viable.

Podríamos haber seleccionado una conexión solo a la tabla 1 (la de la cabecera) y otra conexión para la tabla 2 (la de detalle), ver Ilustración 12.

En este caso vamos a crear una nueva conexión igual que la anterior para trabajar con el detalle y nos basaremos igualmente en el objeto Page001 de la Ilustración 12, pero como hemos señalado podríamos hacer seleccionado el contenido específico que nos interesa, en

este caso Table002, pero con el fin de practicar las tareas de limpieza y transformación lo haremos algo más largo.

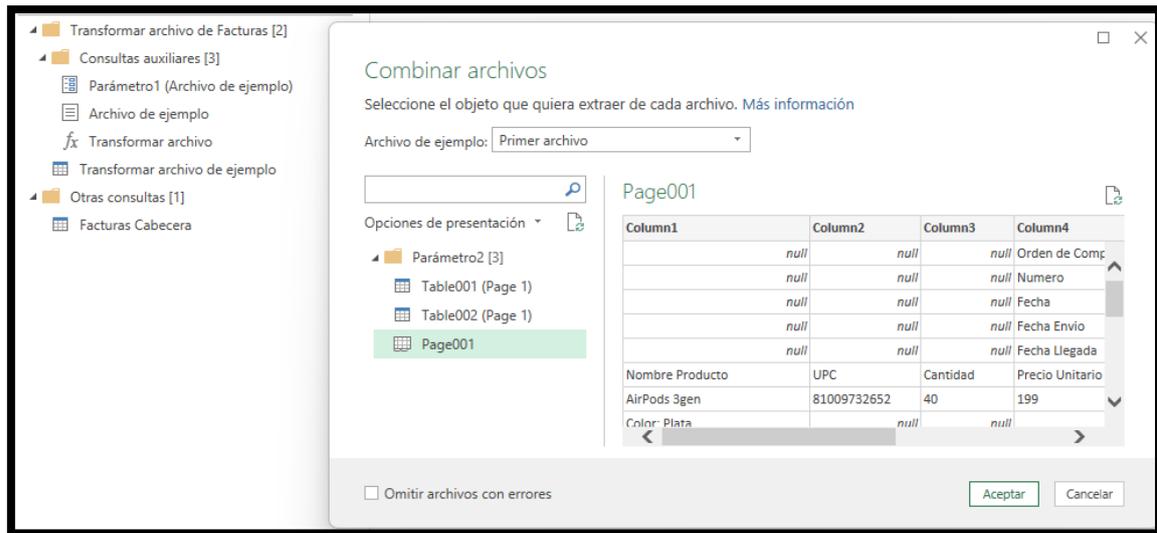


Ilustración 15

El resultado es que disponemos de 2 conexiones a la misma carpeta “Facturas” con sus respectivas estructuras y archivos de ejemplo vinculados y sobre los que se definirán las plantillas que serán replicadas en Facturas Cabeceras y Facturas Detalle. Así el archivo de ejemplo C1 afectará la cabeceras y el C2 a Detalles.

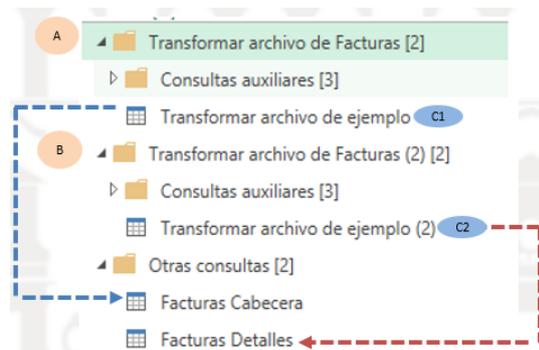


Ilustración 16

3.3 Creando el parámetro de conexión y aplicando

Es recomendable contar con un parámetro de conexión y aplicarlo a las distintas conexiones de carpetas que hayamos creado ya que esto permite mejorar significativamente la flexibilidad en la actualización del origen de datos al cambiar fácilmente las fuentes (como rutas de archivos o nombres de servidores) sin tener que modificar cada consulta individualmente lo que hace que el mantenimiento sea más sencillo.

En la siguiente ilustración hemos creado el parámetro de conexión denominado “ConexiónCarpeta” y que contiene la ruta a la fuente de datos (carpeta que nos interesa), así una vez creado el parámetro tendremos que cambiar el origen de las 2 conexiones al citado parámetro, tal y como se muestra en la Ilustración 17, es decir para la primera conexión (azul) tendremos que cambiar el origen por el parámetro “ConexionParamteros” en los elementos Archivo de ejemplo y Facturas Cabecera y para la segunda conexión (rojo) le tendremos que asignar el parámetro en los elementos Archivo de ejemplo (2) y Facturas Detalles.

Consultas [11]

- Transformar archivo de Facturas [2]
- Consultas auxiliares [3]
 - Parámetro1 (Archivo de ejemplo)
 - Archivo de ejemplo
 - Transformar archivo
 - Transformar archivo de ejemplo
- Transformar archivo de Facturas (2) [2]
- Consultas auxiliares [3]
 - Parámetro2 (Archivo de ejemplo (2))
 - Archivo de ejemplo (2)
 - Transformar archivo (2)
 - Transformar archivo de ejemplo (2)
- Otras consultas [3]
 - Facturas Cabecera
 - Facturas Detalles
 - ConexionParametros (C:\Users\joseignaci...

fx = Folder.Files(ConexionParametros)

	Content	A ^B _C Name	A ^B _C Extension
1	Binary	OC-1.pdf	.pdf
2	Binary	OC-2.pdf	.pdf

Carpeta

Ruta a la carpeta

ConexionParametros

Ilustración 17

4 Transformaciones básicas

4.1 Transformación de la consulta “Cabecera” (Facturas Consolidadas v2a ETL Cabecera)

Fichero: Facturas Consolidadas v2a ETL Cabecera

4.1.1 Transformar archivo de ejemplo, objetivo

Vamos a proceder a continuación a aplicar un proceso de limpieza y transformación que permita disponer de la cabecera de todas las facturas y cómo podemos observar en la Ilustración 8 e Ilustración 18), estos se concentran en las primeras filas. Estos ajustes se realizarán sobre el elemento “Transformar archivo de ejemplo” que se convertirá en una plantilla que se replicará a todos los ficheros contenidos en la carpeta conectada (Facturas) y se combinarán y transformarán generando así la consulta “Facturas Cabecera”.

A ^B Column1	A ^B Column2	A ^B Column3	A ^B Column4	A ^B Column5
1	null	null	null	Orden de Compra
2	null	null	null	Numero
3	null	null	null	Fecha
4	null	null	null	Fecha Envío
5	null	null	null	Fecha Llegada
6	Nombre Producto	UPC	Cantidad	Precio Unitario
7	AirPods 3gen	81009732652	40	199
8	Color: Plata	null	null	null
9	HomePod Mini	81009732653	45	99
10	Color: Azul	null	null	null
11	TotalMount Pro Max	81009732654	40	35.95
12	Color: Negro	null	null	null
13	Soporte Magnético Bekin	81009732655	50	34.95
14	Color: Negro	null	null	null
15	Funda Smart Folio	81009732656	100	89
16	Color: Azul Mar	null	null	null
17	Etiqueta para maleta	81009732658	45	449
18	Color: Marrón	null	null	null
19	null	null	null	Subtotal
20	null	null	null	Total
21	Notas	null	null	null
22	null	null	null	Por favor confirmar recepción al correo. Gracias

Ilustración 18

4.1.2 Limpieza de 1º Grado-Técnicas de depuración

Procedemos a aplicar los ajustes necesarios de limpieza y transformación relacionados con el Archivo de ejemplo vinculado a la consulta Facturas Cabeceras, con el objetivo de disponer solo de los datos de la cabecera y eliminar el resto, para lo cual procedemos a conservar las primeras 5 filas (**Conservar filas superiores**).

Conservar filas superiores

Especifique cuántas filas desea conservar.

Número de filas:

Ilustración 19

El resultado de este paso en el que se muestra en la siguiente Ilustración donde hemos aislado los datos de la cabecera del detalle y del pie del documento.

	A ^B Column1	A ^B Column2	A ^B Column3	A ^B Column4	A ^B Column5
1	null	null	null	Orden de Compra	null
2	null	null	null	Numero	176943
3	null	null	null	Fecha	11/04/2020
4	null	null	null	Fecha Envio	28/06/2020
5	null	null	null	Fecha Llegada	10/07/2020

Ilustración 20

Procedemos a **quitar la primera fila** (Quitar filas Superiores) que no nos interesa y también las 4 primeras columnas (Quitar columnas seleccionadas), con lo que el resultado obtenido será el siguiente.

	ABC 123 Column4	ABC 123 Column5
1	Numero	176943
2	Fecha	11/04/2020
3	Fecha Envio	28/06/2020
4	Fecha Llegada	10/07/2020

Ilustración 21

Ahora nos queda llevar a cabo el proceso de transposición (**Transponer**) que lo que hace es convertir la filas como columnas y las columnas como filas para lo cual seleccionamos la primera columna de la tabla que deseamos transponer y procedemos obteniendo como resultado el siguiente.

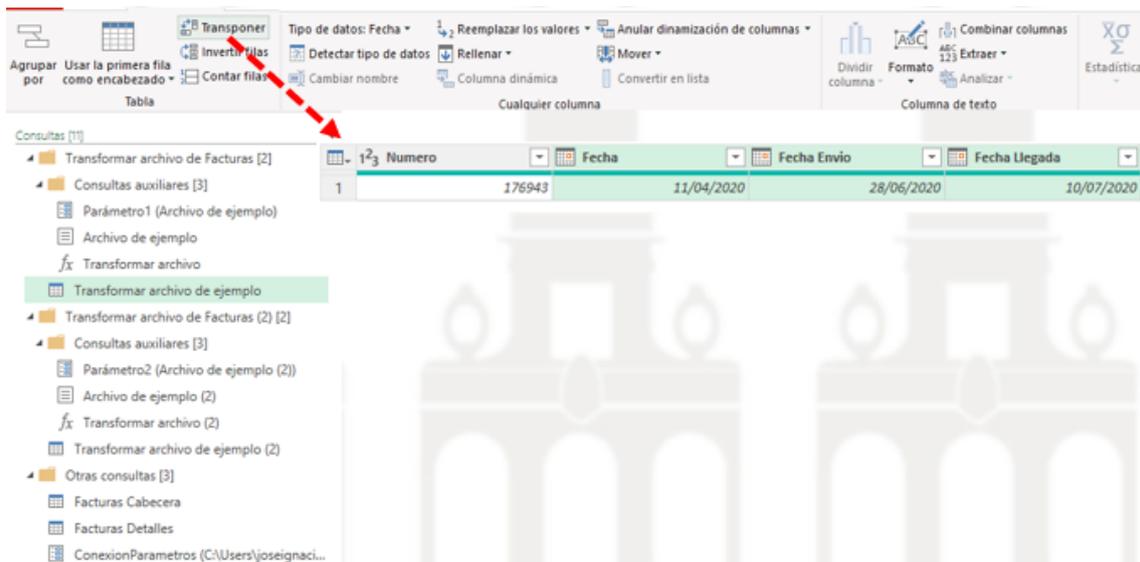


Ilustración 22

Una vez realizada esta transformación realizamos una serie de ajustes básicos finales como, **Encabezados promovidos** y **Tipo cambiado** de las columnas, especialmente el formato número y fecha. El resultado final es que estos ajustes se replican sobre todas los archivos contenidos en la carpeta a la que hemos conectado (Facturas) y se combinan, obteniendo como resultado el siguiente:

	A ^B Source.Name	ABC 123 Numero	ABC 123 Fecha	ABC 123 Fecha Envio	ABC 123 Fecha Llegada
1	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
2	OC-2.pdf	159753	11/04/2020	28/06/2020	10/07/2020

Ilustración 23

4.1.3 Resumen del proceso Limpieza de Datos de 1º Grado, Técnicas de Depuración

En resumen, los ajustes aplicados sobre el fichero “Transformar archivo de ejemplo” son:

- A. Registro del origen de ejemplo y elemento (ver Ilustración 11)
- B. En este apartado hemos aplicado los ajustes básicos como: **Conservar filas, Quitar filas y columnas, transponer, tipo cambiado, encabezados promovidos.**

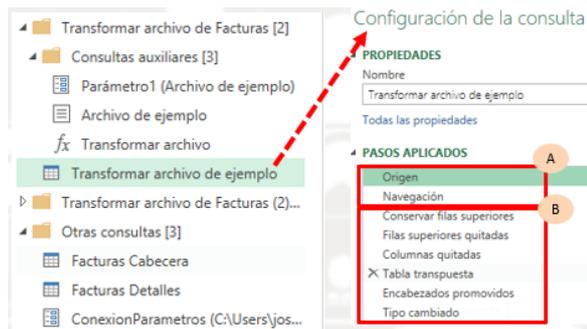


Ilustración 24

4.2 Transformación de la consulta “Detalle” (Facturas Consolidadas v2b ETL Detalle)

Fichero: Facturas Consolidadas v2b ETL Detalle

4.2.1 Transformar archivo de ejemplo (2), objetivo

A continuación, realizaremos el proceso ETL aplicado al area de detalle de la factura, es decir nos interesa obtener los registros correspondientes al detalle de la factura (ver Ilustración 8) que se corresponde al area de B de la Ilustración 25, entre la cabecera (area A) y el pie (area C),

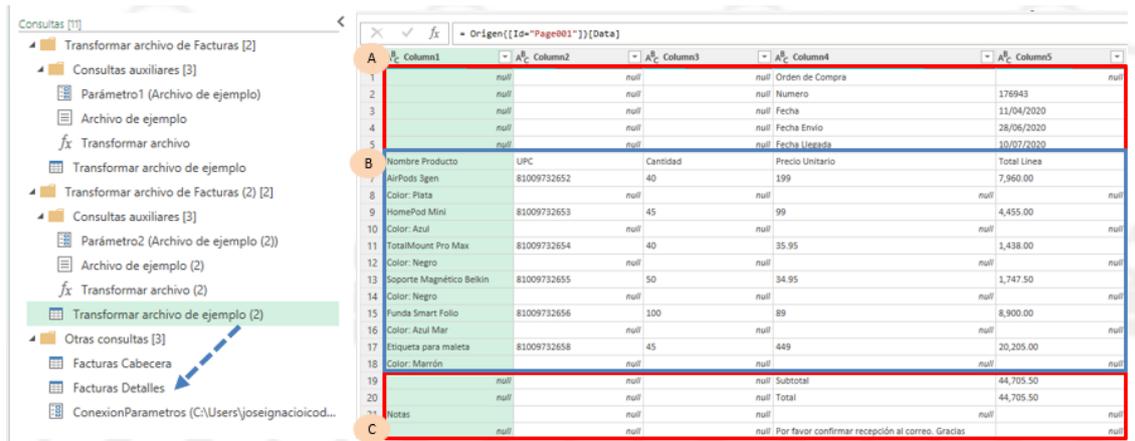


Ilustración 25

Estos ajustes se realizarán sobre el elemento “Transformar archivo de ejemplo (2)” que se convertirá en una plantilla que se replicara a todos los ficheros contenidos en la carpeta conectada (Facturas) y se combinaran y transformaran generando así la consulta “Facturas Detalles”.

Procedemos por tanto a realizar los ajustes de transformación aplicando técnicas de limpieza de primer grado y de segundo grado.

4.2.2 Limpieza de 1º grado (depuración) y Limpieza de 2º grado (Dinamización por transposición)

De esta forma los ajustes básicos comenzaran eliminando

la cabecera de la factura (Quitar filas superiores) y el pie (quitar filas inferiores)

Comenzamos por quitar las 5 filas superiores (**Quitar filas superiores**) que se corresponde con la cabecera la factura. Igualmente, quitaremos las 4 ultimas filas (**Quitar filas inferiores**) que se corresponde con el pie de la factura y con el total y subtotal que no necesitamos

porque estos últimos (total y subtotal) son campos calculados. Así el resultado de este primer proceso de depuración será el siguiente:

	A ^B _C Column1	A ^B _C Column2	A ^B _C Column3	A ^B _C Column4	A ^B _C Column5
1	Nombre Producto	UPC	Cantidad	Precio Unitario	Total Linea
2	AirPods 3gen	81009732652	40	199	7,960.00
3	Color: Plata		null	null	null
4	HomePod Mini	81009732653	45	99	4,455.00
5	Color: Azul		null	null	null
6	TotalMount Pro Max	81009732654	40	35.95	1,438.00
7	Color: Negro		null	null	null
8	Soporte Magnético Belkin	81009732655	50	34.95	1,747.50
9	Color: Negro		null	null	null
10	Funda Smart Folio	81009732656	100	89	8,900.00
11	Color: Azul Mar		null	null	null
12	Etiqueta para maleta	81009732658	45	449	20,205.00
13	Color: Marrón		null	null	null

Ilustración 26

Vamos a proponer los encabezado a la primera fila y cambiar el nombre de alguno de ellos (Usar la primera fila como encabezado).

El siguiente paso es afrontar el problema del color que como podemos observar aparece en una línea secundaria debajo de cada producto (ver Ilustración 8, triangulo rojo) y nos interesa disponer el color en una columna, para ello una posible alternativa seria aplicar **Dividir Columna por delimitador personalizado** (Color:) de la siguiente forma:

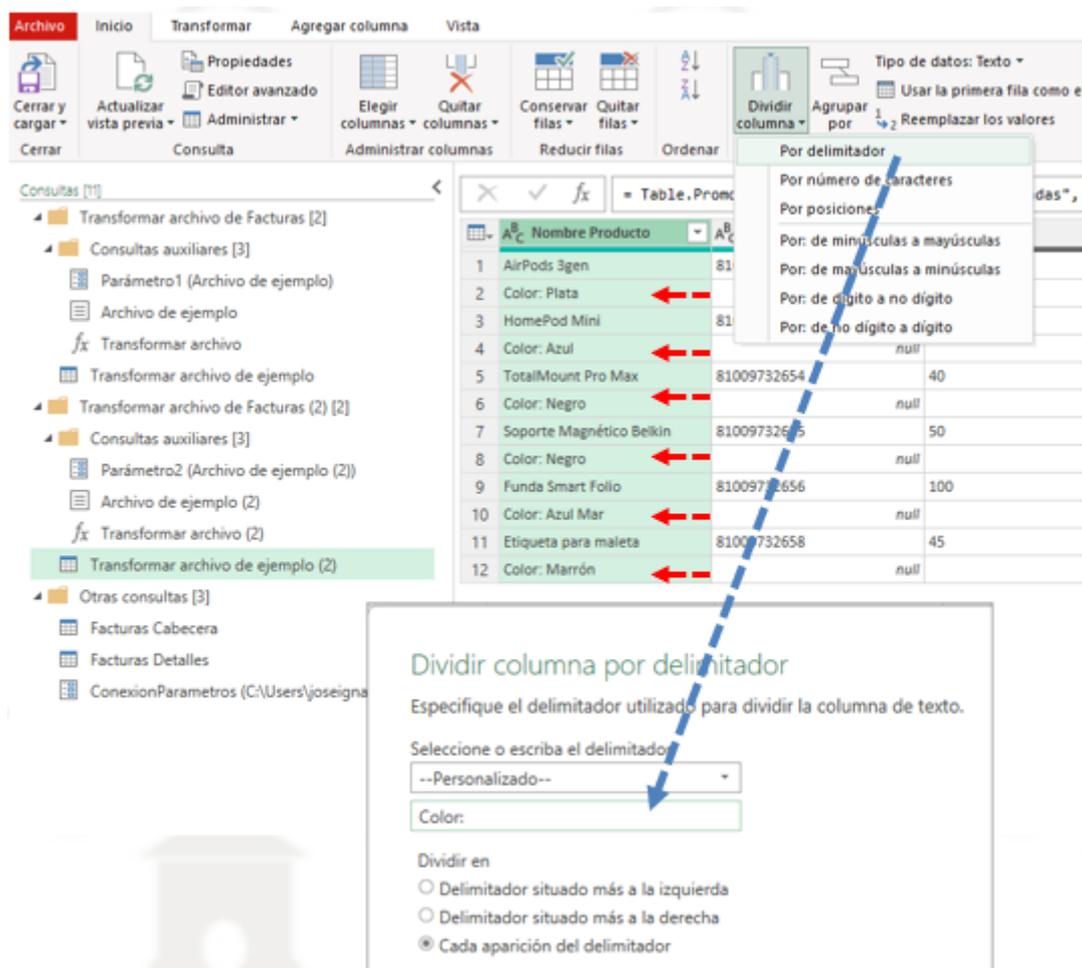


Ilustración 27

El resultado es una nueva columna generada y sobre esta seleccionamos **rellanar hacia arriba**.

A ^B _C Nombre Producto.1	A ^B _C Nombre Producto.2	A ^B _C UPC	A ^B _C Cantidad	A ^B _C Precio Unitario	A ^B _C Total Línea
1	AirPods 3gen	81009732652	40	199	7,960.00
2	Plata	null	null	null	null
3	HomePod Mini	81009732653	45	99	4,455.00
4	Azul	null	null	null	null
5	TotalMount Pro Max	81009732654	40	35.95	1,438.00
6	Negro	null	null	null	null
7	Soporte Magnético Belkin	81009732655	50	34.95	1,747.50
8	Negro	null	null	null	null
9	Funda Smart Folio	81009732656	100	89	8,900.00
10	Azul Mar	null	null	null	null
11	Etiqueta para maleta	81009732658	45	449	20,205.00
12	Marrón	null	null	null	null

Ilustración 28

Otro de los ajustes básicos aplicar sería **filtrar los valores en blanco** del nombre del producto y **cambiar nombres a las columnas** para que sean más identificativos de su contenido.

A ^B _C Producto	A ^B _C Color	A ^B _C Código	A ^B _C Cantidad	A ^B _C Precio Unitario	A ^B _C Total Línea
1	AirPods 3gen	81009732652	40	199	7,960.00
2	HomePod Mini	81009732653	45	99	4,455.00
3	TotalMount Pro Max	81009732654	40	35.95	1,438.00
4	Soporte Magnético Belkin	81009732655	50	34.95	1,747.50
5	Funda Smart Folio	81009732656	100	89	8,900.00
6	Etiqueta para maleta	81009732658	45	449	20,205.00

Ilustración 29

Otra tarea importante en este proceso de limpieza y transformación es verificar el tipo de datos contenidos de las columnas (**Cambiar tipo**) y debemos prestar especial atención al formato numérico, en concreto en nuestro caso en las columnas Precio Unitario y Total Línea donde los valores utilizan el punto como separador decimal y la coma como separador de millar (formato EEUU y Reino Unido) y ajustar el tipo de datos al formato Europeo, esto implica usar la función (**Reemplazar valores**) considerando la siguientes sustituciones y en este orden, para los dos columnas:

1. Quitamos todas las comas en las dos columnas

Ilustración 30

2. Quitadas las comas, reemplazamos los puntos por comas para marcar los decimales

Ilustración 31

El ultimo ajuste por tanto es ya revisar los tipos de campo que contiene las columnas y asi el resultado final del ajuste sobre la tabla Detalle sería el siguiente.

Producto	Color	Codigo	Cantidad	Precio Unitario	Total Linea
1 AirPods 3gen	Plata	81009732652	40	199	7960
2 HomePod Mini	Azul	81009732653	45	99	4455
3 TotalMount Pro Max	Negro	81009732654	40	35,95	1438
4 Soporte Magnético Belkin	Negro	81009732655	50	34,95	1747,5
5 Funda Smart Folio	Azul Mar	81009732656	100	89	8900
6 Etiqueta para maleta	Marrón	81009732658	45	449	20205

Ilustración 32

4.2.3 Resumen del proceso Limpieza de 1º grado y 2º grado

La configuración de la consulta se resume en dos partes como en el caso anterior:

En una primera (A) parte se generan los pasos de forma automática cuando creamos la conexión a la carpeta y seleccionamos el fichero y la columna que contiene los datos.

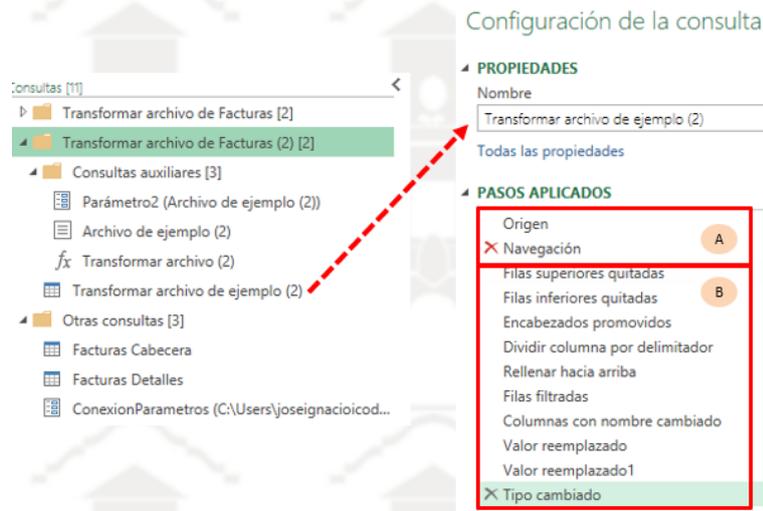


Ilustración 33

En la segunda parte (B) hemos aplicado las transformaciones y depuraciones necesarias usando funciones como: **Quitar filas superiores e inferiores, usar la primera fila como encabezado, Dividir Columna por delimitador personalizado, rellenar hacia arriba, filtrar los valores en blanco, Reemplazar valores, cambiar el nombre a las columnas, reemplazar valores la coma por vacío o blanco (quitar comas), reemplazar valores punto por coma y finalmente ajustamos los tipos de las columnas a texto, número entero y número decimal.**

El resultado final del proceso de limpieza y transformación genera dos tablas con la información consolidada de las facturas correspondiente a la cabecera y el detalle.

Source.Nam	Numer	Fecha	Fecha Envio	Fecha Llegada
OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
OC-2.pdf	159753	11/04/2020	28/06/2020	10/07/2020

Source.N	Producto	Color	Codigo	Cantid	Precio Unitar	Total Linea
OC-1.pdf	AirPods 3gen	Plata	81009732652	40	199,00	7960,00
OC-1.pdf	HomePod Mini	Azul	81009732653	45	99,00	4455,00
OC-1.pdf	TotalMount Pro Max	Negro	81009732654	40	35,95	1438,00
OC-1.pdf	Soporte Magnético Belkin	Negro	81009732655	50	34,95	1747,50
OC-1.pdf	Funda Smart Folio	Azul Mar	81009732656	100	89,00	8900,00
OC-1.pdf	Etiqueta para maleta	Marrón	81009732658	45	449,00	20205,00
OC-2.pdf	Mochila de cuero	Marrón	81893502692	50	35,80	1790,00

Ilustración 34

5 Consulta combinada Cabecera y Detalle (Facturas Consolidadas v3 Combina Cabecera_Detalle)

Fichero: Facturas Consolidadas v3 Combina Cabecera_Detalle

Para finalizar vamos a crear una consulta combinada para lo cual accedemos a la opción **combinar consultas para crear una nueva** y se nos presenta un cuadro de dialogo para su configuración (Ilustración 35).

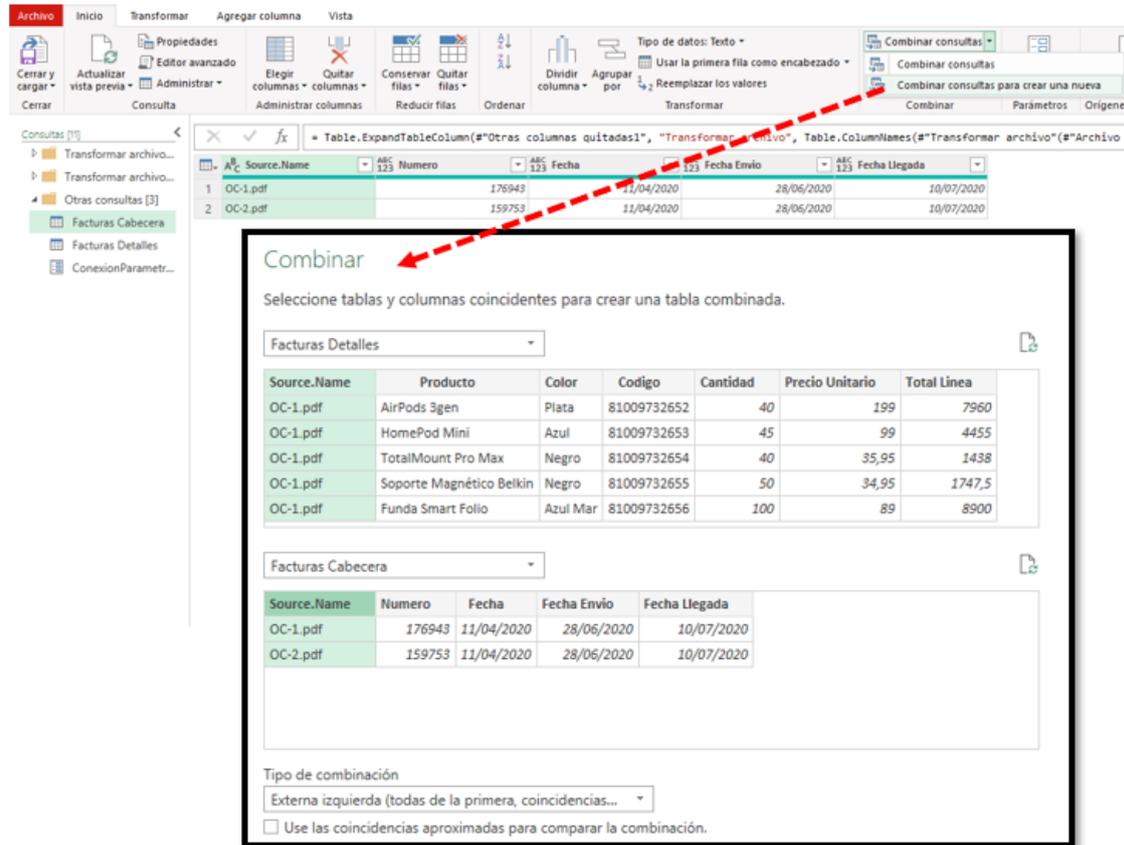


Ilustración 35

La combinación de la nueva consulta se realiza a través de la columna Source.Name, tal y como se muestra en la Ilustración y el resultado será una nueva consulta creada (Combinar 1) añadiendo a Facturas Detalle el contenido de los datos de la Facturas Cabecera a través de una nueva columna generada que denomina con el mismo nombre que la consulta (Facturas Cabecera) tal y como vemos a continuación en la Ilustración 37 y donde podremos desplegar su contenido para seleccionar que elementos deseamos visualizar.

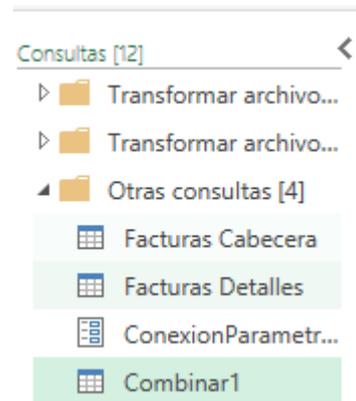


Ilustración 36

	Source.Name	Producto	Color	Codigo	Cantidad	Precio Unitario	Total Linea	Facturas Cabecera
1	OC-1.pdf	AirPods 3gen	Plata	81009732652	40	40	199	7960 Table
2	OC-1.pdf	HomePod Mini	Azul	81009732653	45	45	99	4455 Table
3	OC-1.pdf	TotalMount Pro Max	Negro	81009732654	40	40	35,95	1438 Table
4	OC-1.pdf	Soporte Magnético Belkin	Negro	81009732655	50	50	34,95	1747,5 Table
5	OC-1.pdf	Funda Smart Folio	Azul Mar	81009732656	100	100	89	8900 Table
6	OC-1.pdf	Etiqueta para maleta	Marrón	81009732658	45	45	449	20205 Table
7	OC-2.pdf	Mochila de cuero	Marrón	81893502692	50	50	35,8	1790 Table

Buscar columnas para expandir A Z

Expandir Agregar

(Seleccionar todas las columnas)

Source.Name

Numero

Fecha

Fecha Envio

Fecha Llegada

Usar el nombre de columna original como prefijo

Ilustración 37

Para finalizar, vamos a desplegar los campos que queremos visualizar de la Combinada1 y cambiarle el nombre a esta consulta ademas de hacer los ajustes de cambio de tipo de las columnas.

El resultado final es el mostrado en Ilustración 38 donde por limitaciones de espacio hemos separado las columnas en dos para poder visualizarlas claramente, obteniendo asi el objetivo de consulta combinada que hemos llamado como "Cabecera y Detalle".

Consultas [12]

- Transformar archivo de...
- Transformar archivo de...
- Otras consultas [4]
 - Facturas Cabecera
 - Facturas Detalles
 - ConexionParametros...
 - Cabecera y Detalle

	Source.Name	Producto	Color	Codigo	Cantidad	Precio Unitario	Total Linea	Numero
1	OC-1.pdf	AirPods 3gen	Plata	81009732652	40	40	199	17694
2	OC-1.pdf	HomePod Mini	Azul	81009732653	45	45	99	17694
3	OC-1.pdf	TotalMount Pro Max	Negro	81009732654	40	40	35,95	1438
4	OC-1.pdf	Soporte Magnético Belkin	Negro	81009732655	50	50	34,95	1747,5
5	OC-1.pdf	Funda Smart Folio	Azul Mar	81009732656	100	100	89	8900
6	OC-1.pdf	Etiqueta para maleta	Marrón	81009732658	45	45	449	20205
7	OC-2.pdf	Mochila de cuero	Marrón	81893502692	50	50	35,8	1790

	Source.Name	Numero	Fecha	Fecha Envio	Fecha Llegada
1	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
2	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
3	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
4	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
5	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
6	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020
7	OC-2.pdf	159753	11/04/2020	28/06/2020	10/07/2020

Ilustración 38

6 Resultado final objetivo

El resultado final de todo el proceso ETL de conexión a la carpeta Facturas queda reflejado en nuestro documento de trabajo en tres pestañas correspondientes a las consultas creadas.

	A	B	C	D	E	F
1	Source.Name	Numer	Fecha	Fecha Envio	Fecha Llegada	
2	OC-1.pdf	176943	11/04/2020	28/06/2020	10/07/2020	
3	OC-2.pdf	159753	11/04/2020	28/06/2020	10/07/2020	
4						

Facturas Cabecera

	A	B	C	D	E	F	G
1	Source.Nº	Producto	Color	Codigo	Cantid	Precio Unitar	Total Linea
2	OC-1.pdf	AirPods 3gen	Plata	81009732652	40	199,00	7960,00
3	OC-1.pdf	HomePod Mini	Azul	81009732653	45	99,00	4455,00
4	OC-1.pdf	TotalMount Pro Max	Negro	81009732654	40	35,95	1438,00
5	OC-1.pdf	Soporte Magnético Belkin	Negro	81009732655	50	34,95	1747,50
6	OC-1.pdf	Funda Smart Folio	Azul Mar	81009732656	100	89,00	8900,00
7	OC-1.pdf	Etiqueta para maleta	Marrón	81009732658	45	449,00	20205,00
8	OC-2.pdf	Mochila de cuero	Marrón	81893502692	50	35,80	1790,00
9							

Facturas Detalles

	A	B	C	D	E	F	G	H	I	J	K
1	Source.Name	Producto	Color	Codigo	Cantidad	Precio Unitario	Total Linea	Numero	Fecha	Fecha Envio	Fecha Llegada
2	OC-1.pdf	AirPods 3gen	Plata	81009732652	40	199	7960	176943	11/04/2020	28/06/2020	10/07/2020
3	OC-1.pdf	HomePod Mini	Azul	81009732653	45	99	4455	176943	11/04/2020	28/06/2020	10/07/2020
4	OC-1.pdf	TotalMount Pro Max	Negro	81009732654	40	35,95	1438	176943	11/04/2020	28/06/2020	10/07/2020
5	OC-1.pdf	Soporte Magnético Belkin	Negro	81009732655	50	34,95	1747,5	176943	11/04/2020	28/06/2020	10/07/2020
6	OC-1.pdf	Funda Smart Folio	Azul Mar	81009732656	100	89	8900	176943	11/04/2020	28/06/2020	10/07/2020
7	OC-1.pdf	Etiqueta para maleta	Marrón	81009732658	45	449	20205	176943	11/04/2020	28/06/2020	10/07/2020
8	OC-2.pdf	Mochila de cuero	Marrón	81893502692	50	35,8	1790	159753	11/04/2020	28/06/2020	10/07/2020
9											
10											

Cabecera y Detalle