



# SEMANA DE POWER QUERY

Día 1: Las Acciones Primordiales en Preparación de Datos



Escuela de Inteligencia de Negocios | [ExcelFreeBlog.com](https://www.ExcelFreeBlog.com)



# *Proceso* **ETL**

# Conceptos de Fuego para Trabajar con Power Query

*El Proceso ETL*



## Proceso ETL

Las siglas derivan del inglés: extract, transform and load. Este proceso consiste en llevar datos de múltiples orígenes, para luego hacer la limpieza y transformación necesaria para finalmente dejarla en el destino listos para el consumo y utilización para posterior trabajo en análisis.

Históricamente el proceso ha lucido así:





# *Metodología* **Jerárquica A/0**

# Conceptos de Fuego para Trabajar con Power Query

*Metodología Jerárquica A/O: Acciones, Operaciones y Sub operaciones*



## Metodología Jerárquica A/O

---

La metodología jerárquica de acciones y operaciones ha sido nuestro esfuerzo en categorizar cada actividad que podemos realizar en Power Query, si bien, aún seguimos trabajando en ello, a continuación, veremos de que se trata.

---

# Conceptos de Fuego para Trabajar con Power Query

Metodología Jerárquica A/O: Acciones, Operaciones y Sub operaciones



## LIMPIEZA DE DATOS

Limpieza de Datos  
de Primer Grado:

Técnicas de  
Depuración

- Reducción
  - *Exclusión de Filas*
  - *Quitar Columnas*
  - *Promover Encabezados*
- Formateo de Datos
  - *Transformación a Nivel de Columnas*
  - *Reemplazar Valores*
  - *Dividir Columnas*
  - *Asignar Nombres*
- Transposición
- Rellenar

Limpieza de Datos  
de Segundo Grado:

Técnicas de  
Dinamización por  
Transposición

- Anulación de Dinamización (Unpivot)
- Columna Dinámica (Pivot)

## INTEGRACIÓN DE DATOS

- Anexar
- Combinar
  - *Externa Izquierda*
  - *Externa Derecha*
  - *Externa Completa*
  - *Interna*
  - *Anti Izquierda*
  - *Anti Derecha*
- Agrupar

## ENRIQUECIMIENTO DE DATOS

- Horizontal
- Vertical



# *Topología de* **Formatos de Tabla**

# Topología de Formatos de Tablas

En Búsqueda de Funciones Personalizadas Estándar



## ¿De Qué Va La Topología?

La topología de tablas es un nuevo proyecto que ha surgido mientras se redactaban las páginas del libro El ADN de Power Query, el cual consisten en crear toda una nueva rama en preparación de datos inteligente que se dedique al estudio de las propiedades de los distintos formatos de tablas, sin alterar su significado e información subyacente.

Todo con el fin de crear: *ejemplos, guías, referencias y funciones personalizadas*, etc. para este tipo de situaciones, bien sea CASOS comunes o de frecuencia mínima.

Hasta el momento sólo se ha planteado la idea inicial, sin embargo, los dos capítulos finales del libro El ADN de Power Query y las especializaciones en Power Query buscan empezar a encaminar el proyecto de forma más sólida y formal.





# Topología de Formatos de Tablas

En Búsqueda de Funciones Personalizadas Estándar



## ¿De Qué Va La Topología?

La topología de tablas es un nuevo proyecto que ha surgido mientras se redactaban las páginas del libro El ADN de Power Query, el cual consisten en crear toda una nueva rama en preparación de datos inteligente que se dedique al estudio de las propiedades de los distintos formatos de tablas, sin alterar su significado e información subyacente.

Todo con el fin de crear: *ejemplos, guías, referencias y funciones personalizadas*, etc. para este tipo de situaciones, bien sea CASOS comunes o de frecuencia mínima.

Hasta el momento sólo se ha planteado la idea inicial, sin embargo, los dos capítulos finales del libro El ADN de Power Query y las especializaciones en Power Query buscan empezar a encaminar el proyecto de forma más sólida y formal.



# Topología de Formatos de Tablas

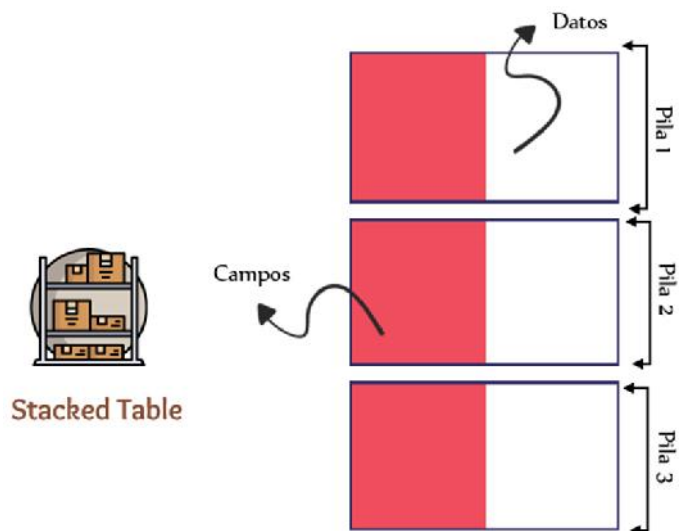
*En Búsqueda de Funciones Personalizadas Estándar*



## Un Componente Inicial

Un componente inicial de la topología de tabla es tener esquemas visuales resumidos de cada tipo de tabla que se vaya encontrando, además, de un único representativo, esto nos permitirá resolver escenarios complejos viviendo en pequeñas partes más sencillas

**Ejemplo:**



# Topología de Formatos de Tablas

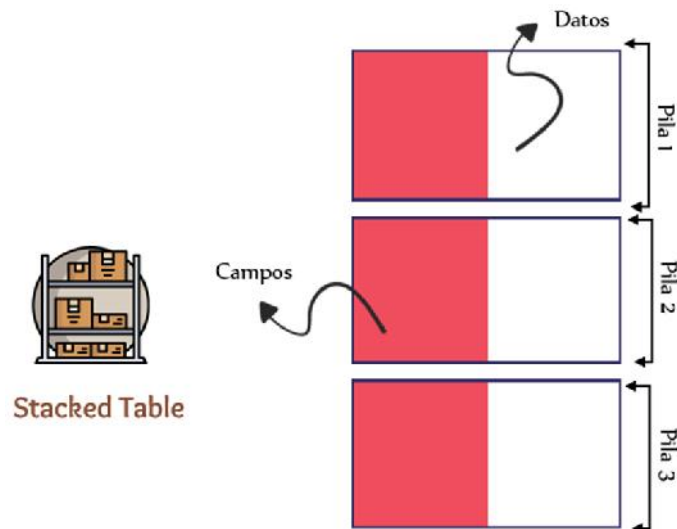
*En Búsqueda de Funciones Personalizadas Estándar*




## Un Componente Inicial

Un componente inicial de la topología de tabla es tener esquemas visuales resumidos de cada tipo de tabla que se vaya encontrando, además, de un único representativo, esto nos permitirá resolver escenarios complejos viviendo en pequeñas partes más sencillas

**Ejemplo:**





# *Power Query*



# La Tecnología de Preparación y Limpieza de Datos

*Conocido también como: Obtener y Transformar*

## POWER QUERY

A.K.A: Obtener y Transformar

Power Query es una tecnología descomunal de alcance infinito, ya que permite: entender, explorar, *preparar, adecuar, limpiar, formatear, integrar, enriquecer, crear, refinar, formular* y hasta *resumir* datos de forma veloz, intuitiva y coherente.

Si bien Power Query tuvo sus inicios y gran parte de su evolución en Excel, sus incontables bondades son tan útiles que se ha propagado a otros programas de Microsoft, concretamente a la fecha de redacción de esta presentación, Power Query se puede encontrar en las siguientes tecnologías: **Microsoft Excel, Power BI Desktop, Power BI Services, Microsoft SQL Server Data Tools, Microsoft SQL Server Analysis Services, Azure Analysis Services, Common Data Services for Apps y Microsoft Flow.**



# Extracción, Transformación y Carga

*El Proceso ETL*

1



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

2



**Transformación (Transform):** Después de conectarte al origen y extraer los datos, si es necesario, tienes que limpiarlos, integrarlos o incluso enriquecerlos para alcanzar un Formato Tabular, todo utilizando diversas técnicas de Limpieza e Integración de Datos.

3



**Cargar (Load):** Con los datos esculpidos en un formato tabular, es momento de cargarlos en algún destino dentro de Excel: En una Hoja del Libro, en el Motor DAX o crear solo la Conexión. Si estas en Power BI existe un solo camino y es el Motor DAX, todo para su posterior análisis.



## Bloque de Trabajo Fundamental: Consultas

*Conocido también como: Obtener y Transformar*

### CONSULTA

A.K.A: Query

Una consulta está formada por una serie de pasos íntimamente relacionados y creados en un orden específico definidos en el propio lenguaje de expresiones de Power Query denominado: Lenguaje M, dichos pasos tienen como objetivo realizar el proceso de *Extracción, Transformación y Carga de datos*, para posterior consumo en herramientas clientes como Microsoft Excel o Microsoft Power BI Desktop.

En términos simplificados pero acertado, se puede interpretar e interiorizar una consulta como un bloque de trabajo en Power Query, es decir, que pueden existir muchos bloques de trabajo (consultas).

## El Proceso ETL





# El Gran Objetivo de la Preparación y Limpieza de Datos

*El formato tabular óptimo*

## El Gran Objetivo

La transformación de datos consta de 3 grandes acciones, la primera de ellas es la limpieza de datos.

Todas las acciones de esta operación tienen un gran objetivo: llegar al formato tabular óptimo

- **Formato Tabular:** Organizado campos y registro (estilo BD)
- **Óptimo:** Que cumpla con todos los detalles para el análisis





# El Gran Objetivo de la Preparación y Limpieza de Datos

*El formato tabular óptimo*

**Formato Tabular Óptimo:** Consiste en tener todos los datos de tal forma que las filas describan un pedacito de información concreto que es soportado por las columnas, de manera simple, consiste en organizar en columnas y filas en el formato clásico de una tabla o como son ordenados en un gestor de bases de datos

## Características del Formato Tabular Óptimo:

- ✓ Columnas con etiquetas.
  - Minimizar espacios, tildes, primera letra en mayúscula
- ✓ Sin totales ni subtotales.
- ✓ Sin agrupaciones a nivel de filas.
- ✓ Sin agrupaciones a nivel de columnas.
- ✓ Sin columnas vacías, de error o no útiles.
- ✓ Sin filas vacías, de error o no útiles.
- ✓ Tipo de dato apropiado.
- ✓ Sólo columnas de interés para el análisis.
- ✓ Sólo registros de interés para el análisis.

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

# ¡Muchas Gracias!

© Escuela de Inteligencia de Negocios

© Excel Free Blog

**Este Material Fue creado por:**

Ing. Miguel A. Caballero Sierra

Ing. Fabian E. Torres Hernández

Por favor, mantenga este material para su estudio personal



Escuela de Inteligencia de Negocios | [ExcelFreeBlog.com](http://ExcelFreeBlog.com)