



Caso 101 Empleados

1 Casos ETL Modelado de Datos.

Jose Ignacio González Gómez

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

www.jggomez.eu

V.21

Ejercicio Basado: **Curso Power BI ¿Qué es Power Query? Curso en Excel y Power BI, Aprende usar Power Query en Minutos. Profesor: Jorge Luis Herrera ENLACE**

Enunciado

Nuestro programa de RRHH nos reporta un informe relacionado con el personal de la empresa, departamento al que pertenece, fecha de contratación y otros datos relacionados, tal y como se puede ver en la siguiente ilustración.

	A	B	C	D	E	F	G	H	I	J	
1											
2			Empleados actualizados a la fecha								
3											
4											
5			Área	ID	Nombre	Título profesional	Nivel organizacional	Estado civil	Género	Fecha de contratación	Bono (USD)
6			Administración	295847284	ken.sánchez bonilla	Chief Executive Officer	0 S		M	14/01/2009	0
7				245797967	terri.duffy	Vice President of Engineering	1 S		F	31/01/2008	2000
8				509647174	roberto.tamburello	Engineering Manager	2 M		M	11/11/2007	2000
9				245797967	terri.duffy	Vice President of Engineering	1 S		F	31/01/2008	0
10				112457891	rob.walters	Senior Tool Designer	3 S		M	05/12/2007	3500
11			Total área:	5							
12											
13			Área	ID	Nombre	Título profesional	Nivel organizacional	Estado civil	Género	Fecha de contratación	Bono (USD)
14			Contabilidad	695256908	gail.ericson	Design Engineer	0 M		F	06/01/2008	10000
15				998320692	jossef.goldberg	Design Engineer	3 M		M	24/01/2008	8000
16				134969118	dylan.miller	Research and Development M	3 M		M	08/02/2009	8000
17				811994146	diane.margheim	Research and Development En	4 S		F	29/12/2008	3500
18				658797903	gigi.matthew	Research and Development En	4 M		F	16/01/2009	10000
19				879342154	michael.raheem	Research and Development M	4 M		M	03/05/2009	8000
20				974026903	ovidiu.craclium	Senior Tool Designer	3 S		M	05/12/2010	3500
21				480168528	thierry.d'hers	Tool Designer	4 M		M	11/12/2007	8000
22				486228782	janice.galvin	Tool Designer	4 M		F	23/12/2010	10000
23				42487730	michael.sullivan	Senior Design Engineer	3 S		M	08/12/2010	3500
24				56920285	sharon.salavaria	Design Engineer	3 M		F	18/01/2011	10000
25			Total área:	11							

Se puede observar que la información se encuentra no estructurada y no puede ser utilizada directamente para la elaboración de análisis y estudios basados en tablas y gráficos dinámicos, Power BI y herramientas similares.

Se pide

Es por tanto necesario llevar a cabo un proceso ETL con el editor de consultas Power Query para convertir esta fuente en una base normalizada, como la siguiente

	A	B	C	D	E	F	G	H	I	J
1	Área	ID	Nombre	Apellido	Título profesional	Nivel organizacional	Estado civil	Género	Fecha de contratación	Bono (USD)
2	Administración	295847284	Ken	Sánchez Bor	Chief Executive Offi	0 S		M	14/01/2009	0
3	Administración	245797967	Terri	Duffy	Vice President of E	1 S		F	31/01/2008	2000
4	Administración	509647174	Roberto	Tamburello	Engineering Manag	2 M		M	11/11/2007	2000
5	Administración	245797967	Terri	Duffy	Vice President of E	1 S		F	31/01/2008	0
6	Administración	112457891	Rob	Walters	Senior Tool Design	3 S		M	05/12/2007	3500
7	Contabilidad	695256908	Gail	Erickson	Design Engineer	0 M		F	06/01/2008	10000
8	Contabilidad	998320692	Jossef	Goldberg	Design Engineer	3 M		M	24/01/2008	8000
9	Contabilidad	134969118	Dylan	Miller	Research and Deve	3 M		M	08/02/2009	8000
10	Contabilidad	811994146	Diane	Margheim	Research and Deve	4 S		F	29/12/2008	3500
11	Contabilidad	658797903	Gigi	Matthew	Research and Deve	4 M		F	16/01/2009	10000
12	Contabilidad	879342154	Michael	Raheem	Research and Deve	4 M		M	03/05/2009	8000

Para ello crearemos un nuevo fichero Excel y a través de PQ crearemos una conexión a la fuente original y comenzaremos el proceso de normalización.

Como orientación, los pasos seguidos en este proceso ETL son, previamente creamos una nueva hoja de calculo sobre la que operaremos estos ajustes

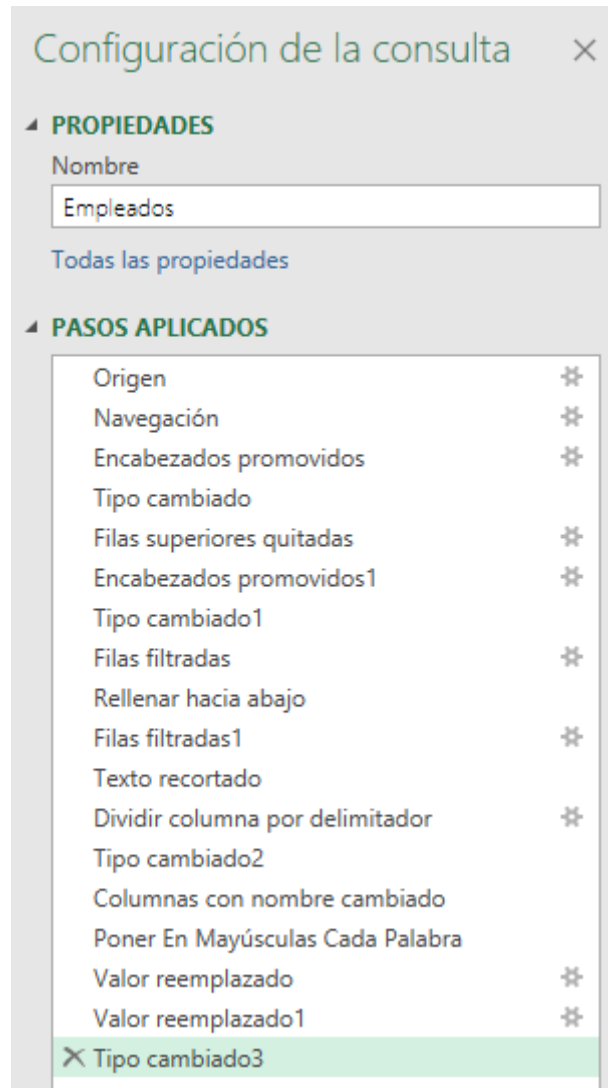
1. FASE CREAR CONEXIÓN – CONSULTA. Así el primer paso es crear una conexión a la fuente a través de Nueva Consulta
2. FASE TRANSFORMACION y LIMPIEZA

- 2.1. Vamos a quitar las primeras filas que no necesitamos y situados sobre A3 seleccionamos la opción "Quitar Filas Superiores"
- 2.2. "Encabezados Promovidos", opción Usar la Primera fila como encabezado
- 2.3. Aplicamos filtros para no disponer de determinadas filas, en concreto en la columna área no nos interesa disponer de aquellas filas que contienen la palabra "Area" así como de "Total área:" y para ello accedemos a la opción filtro para desmarcarlas y disponer de ellas.

	A ^B C Área	ABC ID	A ^B C Nombre	A ^B C Titul profesional
1	Administración	295847284	ken.sánchez bonilla	Chief Executive Officer
2	null	245797967	terri.duffy	Vice President of Engineering
3	null	509647174	roberto.tamburello	Engineering Manager
4	null	245797967	terri.duffy	Vice President of Engineering
5	null	112457891	rob.walters	Senior Tool Designer
6	null	null	null	null
7	Contabilidad	695256908	gail.erickson	Design Engineer
8	null	998320692	jossef.goldberg	Design Engineer
9	null	134969118	dylan.miller	Research and Development Manager

- 2.4. Ahora nos interesa usar la opción rellenar de para que debajo de cada área me rellene los campos que contiene null, esto lo hacemos con la opción "Rellenar Abajo" que accedemos usando el botón derecho sobre la columna deseada y nos da acceso entre otras opciones a esa.
- 2.5. Ocultamos ahora aquellas filas que tienen valores null, por ejemplo, filtrando por ID como hicimos anteriormente y deseleccionándolo.
- 2.6. Separar nombre y apellido. Vemos que en la columna Nombre aparecen nombre y apellido y nos interesa separarla. Pero previamente tenemos que hacer un ajuste:
 - 2.6.1. Limpieza. Eliminar los registros que tienen espacios al comienzo, para lo cual necesitamos usar la opción "Recortar" que accedemos con el botón derecho Transformar-Recortar y así eliminamos los espacios vacíos al comienzo del campo y al finalizar el campo.
 - 2.6.2. Una vez hecho este ajuste procederemos a separar nombre y apellidos a través de dividir columnas por delimitador e igualmente accedemos seleccionando la columna y con botón derecho seleccionamos esa opción. Como podemos observar en este caso el delimitador de nombre y apellidos es el punto.
 - 2.6.3. Nos ha creado dos columnas y le cambiaremos su identificación por Nombre y Apellidos.
 - 2.6.4. Para mejorar esto vamos a cambiar a mayúscula cada palabra de las columnas anteriores y para ello seleccionamos las dos nuevas columnas "Nombre" y "Apellidos" y con botón derecho accedemos a la Transformar "Poner en mayúscula".
- 2.7. Fijarnos que en la columna nivel organización es una columna numérica y nos interesa cambiar null por 0 y para ello accedemos a la opción "Reemplazar Valores" con el click derecho seleccionado de la columna deseada y reemplazamos el valor null por 0.
- 2.8. Igualmente procederemos con la columna Bonos reemplazaremos null por 0.
- 2.9. Revisamos y cambiamos el tipo de dato que tiene cada columna para ello simplemente con el icono de la izquierda del nombre del campo podemos hacerlo. Así por ejemplo ID podemos cambiarlo a entero, igual que nivel organizacional. A fecha lo cambiamos a tipo fecha

Resumen de los pasos seguidos en el proceso ETL



Conceptos y técnicas requeridas

Los conceptos y técnicas para aplicar con Power Query relacionadas con el proceso ETL son:

Creación de una consulta, conexión a un fichero fuente no normalizado.

Uso de las opciones “Quitar Filas Superiores“ y “Encabezados Promovidos” opción Usar la Primera fila como encabezado

Aplicamos filtros para no disponer de determinadas filas, por ejemplo que contiene null en una determinada columna.

Uso de la opción “Rellenar Abajo” que accedemos usando el botón derecho sobre la columna deseada y nos da acceso entre otras opciones a esa.

Limpieza. Eliminar los registros que tienen espacios al comienzo y o al final, usando la opción Transformar-Recortar.

Separar nombre y apellido, con la opción “Dividir Columnas por delimitador”

Uso de “Reemplazar Valores”, por ejemplo, cambiar null por 0

Cambiar tipo de datos de las columnas a través del icono izquierdo