

Crear un modelo eficiente en cuanto a memoria





Miguel Ángel Franco García

Miguel Angel Franco Garcia

22-5-2019

Crear un modelo de datos eficiente en cuanto a memoria con Excel y Power Pivot

En Excel 2013 o versiones posteriores, puede crear modelos de datos que contiene millones de filas y, a continuación, realizar análisis de datos con estos modelos. Modelos de datos pueden crearse con o sin Power Pivot para admitir cualquier número de tablas dinámicas, gráficos y visualizaciones de Power View en el mismo libro.

Las características de Power Pivot que se introdujo en Excel 2013 y modelado de datos también se aplican a Excel 2016.

Aunque se pueden crear enormes modelos de datos en Excel de forma sencilla, es recomendable no hacerlo por diferentes motivos.

- En primer lugar, los modelos de gran tamaño que contienen muchas tablas y columnas resultan excesivos para la mayoría de los análisis y constituyen listas de campos difíciles de manipular.
- En segundo lugar, los modelos grandes consumen memoria, lo que afecta de forma negativa a otras aplicaciones e informes que comparten los mismos recursos del sistema.
- Por último, en Office 365, tanto SharePoint Online como Excel Web App limitan el tamaño de un archivo de Excel a 10 MB. En los modelos de datos de libros que contienen millones de filas, el límite de 10 MB se alcanzará rápidamente. Consulte Especificación y límites de modelos de datos.

En este artículo, le explico cómo crear un modelo de construcción firme con el que sea fácil trabajar y que emplee menos memoria. Si dedica el tiempo necesario para conocer las prácticas recomendadas para el diseño de modelos eficientes, su trabajo se verá recompensado cuando cree y use cualquier modelo en el futuro, con independencia de si lo visualiza en Excel 2013, en Office 365 SharePoint Online, en Office Online Server o en SharePoint 2013.

También puede ejecutar el optimizador de tamaños. Analiza el libro de Excel y si es posible, los comprime aún más.

Optimizador de tamaños

Workbook Size Optimizer

<i>Important!</i> Selectir	ng a language below will dynamically change the	e complete page content to that language.
Language:	English	Download

Elija la versión a instalar. Personalmente, instale la primera opción.

Choose the download you want

File Name	Size
Workbook Size Optimizer_amd64_windows_8_server_2012.msi	3.2 MB
EULA - MS Workbook Size Optimizer.docx	42 KB
Workbook Size Optimizer_amd64_Windows_7_Server_2008_and_below.msi	3.2 MB
Workbook Size Optimizer_x86_Windows_7_Server_2008_and_below.msi	3.2 MB
Workbook Size Optimizer_x86_windows_8_server_2012.msi	3.2 MB

El optimizador de tamaño de libro de trabajo inspecciona la composición del modelo de datos dentro de su libro de trabajo de Power Pivot o Power View, ve si los datos pueden ocupar menos espacio y, si es posible, permite una mejor compresión, sobre todo si estos datos provienen de fuentes de datos externas. La mejor compresión de tamaño se puede lograr para libros de trabajo basados en bases de datos de SQL Server y hay algunos trucos que podemos hacer para otras fuentes de datos SQL también. El optimizador se instalará como **complemento** para sobresalir y le proporcionará un buen asistente para comprimir mejor el tamaño de su libro

de trabajo. Usando el optimizador, a menudo puede obtener más de 1,000,000 de conjuntos de datos de filas en un libro de menos de 10 MB, compartirlo en SharePoint Online e interactuar con él utilizando la aplicación web de Excel en cualquier navegador.

La instalación es simplemente descargar, y, ejecutar.

Si dentro de Excel, vamos a **Opciones – Complementos – Administrar – Complementos COM** veremos el optimizador.

complementos con	*1	
omplementos <u>d</u> ispo	nibles:	Aceptar
Dax Studio Excel Inquire Microsoft Data St	Add-In reamer for Excel	Cancelar
Microsoft Power Microsoft Power	Map for Excel Pivot for Excel	<u>A</u> gregar
Workbook Size C	ptimizer	Quitar
bicación:	file:///C:/Program Files/DAX Studio/DaxStudio.vsto	plvstolocal
	carga: Cargar al inicio	

Seguimos con la creación del modelo de manera eficiente.

Los modelos de datos de Excel usan el motor de análisis en memoria para almacenar datos en la memoria. El motor implementa potentes técnicas de compresión para reducir los requisitos de almacenamiento, lo que contrae el conjunto de resultados hasta que ocupa una fracción de su tamaño original.

Como promedio, los modelos de datos suelen ser entre 7 y 10 veces más pequeños que los mismos datos en el punto de origen. Por ejemplo, si va a importar 7 MB de datos de una base de datos de SQL Server, el modelo de datos podría ocupar 1 MB o menos en Excel. El grado de compresión conseguido en realidad depende principalmente del número de valores únicos de cada columna (cuantos más valores únicos haya, más memoria será necesaria para almacenarlos).

¿Por qué estamos hablando acerca de compresión y valores únicos? Porque, para crear un modelo eficiente que minimice el uso de la memoria, la clave reside en maximizar la compresión. La forma más sencilla de conseguirlo consiste en deshacerse de las columnas que no necesite, especialmente, si dichas columnas incluyen una gran cantidad de valores únicos.

No hay nada mejor para consumir poca memoria que las columnas inexistentes, la columna más eficiente en cuanto a memoria es aquella que nunca llegó a importarse. Si desea crear un modelo eficiente, mire cada columna y pregúntese si contribuye positivamente al análisis que desea realizar. Si no lo hace o no está seguro, no la incluya. Posteriormente, puede agregar columnas nuevas si las necesita.

Dos ejemplos de columnas que siempre se deben excluir, el *primer ejemplo* está relacionado con los datos procedentes de un almacén de datos. En los almacenes de datos, es frecuente encontrar instrumentos de procesos de **ETL** que cargan y actualizan datos en el almacén. Al cargar datos, se crean columnas como **"Fecha de creación**", **"Fecha de actualización**" y **"Ejecución de ETL**". Ninguna de estas columnas es necesaria en el modelo, por lo que no deben seleccionarse al importar datos.

El <u>segundo ejemplo</u> implica la omisión de la columna de cable principal al importar una tabla de hechos.

Muchas tablas, incluidas las **tablas de hechos**, tienen claves principales. Para la mayoría de las tablas, como las que contienen datos sobre clientes, empleados o ventas, es probable que desee la clave principal de la tabla para crear con ella relaciones en el modelo.

Las **tablas de hechos** son diferentes. En estas, se utiliza la clave principal para identificar cada fila de forma única. Aunque es necesaria para la **normalización**, resulta menos útil en los modelos de datos en los que solamente desee usar dichas columnas para llevar a cabo análisis o establecer relaciones de tablas. Por este motivo, al llevar a cabo la importación desde una tabla de hechos, no incluya su clave principal. Las claves principales en las tablas de hechos consumen una gran cantidad de espacio en el modelo y no ofrecen ningún beneficio, ya que no se pueden usar para crear relaciones.

En los almacenes de datos y las bases de datos multidimensionales, se suele hacer referencia a la tablas de gran tamaño compuestas principalmente por datos numéricos como "**tablas de hechos**". Las tablas de hechos suelen incluir datos sobre transacciones o rendimiento comercial, como puntos de datos sobre costes y ventas agregados y adaptados a las unidades organizativas, productos, segmentos de mercado, regiones geográficas, etc. Todas las columnas de las tablas de hechos que contienen datos empresariales o que se pueden usar para hacer referencias cruzadas a los datos almacenados en otras tablas deben estar incluidas en el modelo para que admitan el análisis de datos. La columna que desea excluir es la columna de **clave principal** de la tabla de hechos, que consiste en una serie de valores únicos que solamente existe en la tabla de hechos. Debido al inmenso tamaño de las tablas de hechos, algunos de los principales beneficios de la eficiencia del modelo proceden de la exclusión de filas o columnas de las tablas de hechos.

Los modelos eficientes solamente contienen las columnas que son realmente necesarias en el libro. Si quiere controlar las columnas que se incluyen en el modelo, deberá usar el **Asistente para la importación de tablas en el complemento Power Pivot** para importar los datos en lugar del cuadro de diálogo "**Importar datos**" de Excel.

Al iniciar el Asistente para la importación de tablas, seleccione las tablas que desee importar.

Asistente para la importación de tablas

? ×

Seleccione las tablas y vistas

Seleccione las tablas y vistas de las que desea importar datos.

Servidor: dani-srv

Base de datos: ContosoRetailDW

Tablas y vistas:

	Tabla de origen	Esquema	Nombre descriptivo	Detalles de filtro	^
] 💷	fact_100000	dbo			
] 📖	fact_1000000	dbo			
	fact_250000	dbo			
	fact_500000	dbo			
] 💷	FactExchangeRate	dbo	-		
] 📼	FactInventory	dbo			
	FactITMachine	dbo			
	FactITSLA	dbo			
	FactOnlineSales	dbo	-		
v 🔳	FactSales	dbo	Ventas reales		
	FactSalesQuota	dbo			
	FactStrategyPlan	dbo			
	sysdiagrams	dbo			
					Y
			ione las tablas relaciona	das Vista previa y	
		< Atrás	Siguiente >	Finalizar	ar

Miguel Ángel Franco García. 22/05/2019 Página 7 | 12

Para cada tabla, puede hacer clic en el botón de **vista previa y filtro** y seleccionar las partes de la tabla que realmente necesite. Le recomendamos que anule la selección de todas las columnas en primer lugar y después marque las columnas que desee una vez que valore si son necesarias para el análisis.

Use la casilla de para la columna c	verific on el	cación para selecc fin de seleccionar	ionar las columnas los valores que se	específicas. Para f deben incluir.	filtrar los datos en una	columna, use la flecha de	splegable		
Nombre de la tabla: FactSales									
🔳 🔄 Venta		🗹 Fec 💌	🗹 cana 💽	🗹 Alma 💌	🗹 Produc 💽	🗹 Promoció 💽	🔲 Moned 💽		
	1	02/01/2007	1	209	956	10	1		
	2	12/02/2007	4	308	766	2	1		
	3	24/01/2008	1	156	1175	11	1		
	4	13/01/2008	2	306	1429	10	1		
	5	22/01/2008	2	306	1133	10	1		
	6	02/07/2007	3	200	2365	3	1		
	7	19/11/2007	4	310	1016	5	1		
	8	10/04/2008	2	307	138	15	1		
	9	14/07/2008	2	199	1731	12	1		
	10	10/04/2009	4	310	497	24	1		
	11	04/02/2007	2	199	1825	2	1		
	12	29/04/2007	1	119	543	1	1		
	13	25/07/2007	1	171	739	3	1		
(· · ·					>		

¿Se pueden filtrar solamente las filas necesarias?

Muchas de las tablas de bases de datos corporativas y almacenes de datos contienen datos históricos acumulados durante largos períodos de tiempo. Asimismo, es posible que descubra que las tablas que le interesan contienen información sobre áreas de negocio no necesaria para el análisis específico.

Al usar el **Asistente para la importación de tablas**, puede filtrar los datos históricos o no relacionados, lo que le ahorraría una gran cantidad de espacio en el modelo. En la siguiente imagen, se usa un filtro de datos para recuperar únicamente las filas que contienen datos para el año en curso, excluyendo los datos históricos innecesarios.

vlostrar filas donde "Uzņēmu	m s';						-	and the second s			
es igual que	v	01/06/2012		8-1	Asistent	te para la impo	rtación de tablas ?				
@Y 00					Asistente para la importación de tablas ?						
		27/08/2012			nnse especificas. Para filtrar los datos en una columna, use la flecha desplegable e se deben incluir.						
			 agosto de 2012 lu. ma. mi. ju. vi. sá. 29 30 31 1 2 3 5 6 7 8 0 10 	do. 4	🔝 🗹 Alma	💷 💌 Prode	uc 💽 🔽 Promoció	. 💽 🔝 Moned.	E A		
		100	12 13 14 15 16 17	18	1	209	956	10	1		
			19 20 21 22 23 24	25	4	308	766	2	1		
			2 3 4 5 6 7	8	1	156	1175	11	1		
			Today: 27/08/20	12	2	306	1429	10	1		
			5 22/01/2008		2	306	1133	10	1		
			6 02/07/2007		3	200	2365	3	1		
			7 19/11/2007		4	310	1016	5	1		
			8 10/04/2008		2	307	138	15	1		
			9 14/07/2008		2	199	1731	12	1		
			10 10/04/2009		4	310	497	24	1		
			11 04/02/2007		2	199	1825	2	1		
			12 29/04/2007		1	119	543	1	1		
			13 25/07/2007		1	171	739	3	1		

¿Qué ocurre si se necesita la columna? ¿Se puede hacer algo más para reducir el espacio consumido?

Hay algunas técnicas adicionales que puede aplicar para convertir una columna en un candidato más adecuado para la compresión.

Modificación de columnas DateTime

En muchos casos, las columnas **Date time** ocupan una gran cantidad de espacio. Por suerte, se pueden reducir los requisitos de almacenamiento para este tipo de datos de diferentes formas. Las técnicas variarán en función de cómo utilice la columna.

Las columnas **Date time** incluyen una parte de fecha y una hora. Cuando se pregunte si necesita una columna, hágase la misma pregunta varias veces para las columnas **Date time**:

- ¿Necesito la misma parte?
- ¿Necesito la misma parte a nivel de horas, minutos, segundos y milisegundos?
- ¿Tengo varias columnas **Date time** porque deseo calcular la diferencia entre ellas o solamente para agregar la fecha por año, mes, trimestre, etc.?

La respuesta a cada una de estas preguntas determinará sus opciones para abordar la columna Date time.

Uso de medidas calculadas de DAX en lugar de columnas

Si ya ha trabajado con el lenguaje de expresión de DAX, posible que sepa que las columnas calculadas se usan para derivar columnas nuevas basándose en otra columna del modelo. Aunque las medidas calculadas se definen una vez en el modelo, solamente se evalúan al usarse en una tabla dinámica u otro informe.

Una técnica para ahorrar memoria consiste en sustituir las columnas normales o calculadas por medidas calculadas. Los ejemplos clásicos son **Precio**, **Cantidad** y **Total**. Si tiene las tres, puede ahorrar espacio si mantiene solo dos y calcula la tercera con DAX.

¿Qué dos columnas debe conservar?

En el ejemplo anterior, conserve **Cantidad** y Precio. Estas dos contienen menos valores que **Total**. Para calcular **Total**, agregue una medida calculada como la siguiente:

"TotalVentas: =sumx ('Datos', 'Datos'[Precio]* 'Datos'[Cantidad])"

Las columnas calculadas son iguales que las consultas normales en el sentido de que ambas ocupan espacio en el modelo. Sin embargo, las medidas calculadas se obtienen sobre la marcha y no ocupan espacio.



Miguel Angel Franco Garcia Certificado en Excel básico, medio, avanzado, análisis de datos, y, Power BI



Miguel Ángel Franco García. 22/05/2019 Página 12|12