

Massive Excel Spreadsheets – How to Reduce the Size of Pivot Table Workbooks

Since we've moved to excel 2010 with its expanded data range the tendency has been to create ever more detailed reports. Where once we summarized data we can now leave in every last bit of detail.

Detailed reports always need summaries which is where the pivot table comes in handy. So now we have huge data sheets and a load of number crunching pivot tables to make sense of it all. It is not a huge surprise that our spreadsheets have ballooned in size.

At month end after processing the SLAM data for multiple trusts I can easily end up with spreadsheets in excess of 40Mb which is a big problem, not only do these take ages to process and take up more than their fair share of storage space but I struggle to distribute them as they won't go through the email system.

That's a big problem and we need a solution which will cut the size of our pivot table spreadsheets.

Pivot Cache

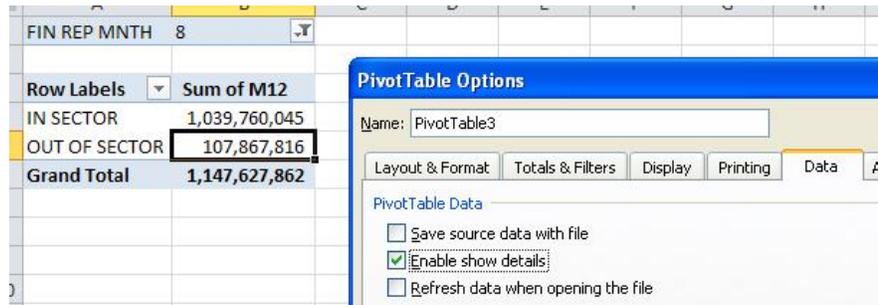
It helps to understand how the problem of large file sizes occurs when using pivot tables

Whenever you create a PivotTable, Microsoft Excel creates a duplicate, hidden copy of the source data. This is known as the pivot cache and enables Excel to quickly calculate and aggregate data without interfering with the original data. It optimizes processing time but it also increases the size of the spreadsheet as it effectively doubles up the amount of data stored with each spreadsheet. If you create multiple pivot tables from the same data source you can end up doubling, tripling, quadrupling etc etc according to the number of additional caches produced.

There are basically two different options for dealing with the pivot cache to reduce the size of pivot table reports, you can either force Excel not to create the Pivot Cache or use the Pivot Cache but optimize by using the same cache for each pivot and removing the original data sheet to avoid duplication.

Store Excel Pivot Table without the Pivot Cache

You can prevent Excel from creating an additional cache of data at the pivot table wizard stage or after you've created the pivot table by going into the pivot table options screen and deselecting the checkbox "Save source data with file".



This doesn't sever the link with the data source so you can still refresh and reorganize your pivot but it does increase the time taken to refresh.

This method does decrease the size of the file but as you can see from the conclusion below that it is not the most efficient way of reducing file sizes.

Store Excel Pivot Table with the Pivot Cache but optimize by:

1. Ensure all pivots use same cache

In older versions of excel it was very easy to build multiple pivot tables from the same range of data but with each pivot creating its own cache and therefore expanding the file size. In Excel 2010 this seems to have been improved and it is much harder to build multiple caches. It is still worth knowing how to avoid this and force the pivot to use the same cache of data.

When I create multiple pivot tables I always copy and paste from another pivot table rather recreating from scratch. This forces it to use the same cache and ensures that any calculated fields, grouping etc. are available across all pivots. This way each additional pivot is practically free in terms of additional file size.

2. Delete the data sheet



If you are allowing the pivot table to store the cache there is no reason to also show a data tab in your workbook as you now have two copies of the same thing.

If you delete the data tab only the cache will be saved and you will save the space associated with the original data table. It is particularly efficient as the cache does not store calculations or formatting but you can no longer refresh the table.

How to retrieve data from pivot

If you did need to retrieve the driving data and display on a single sheet just double click on the Grand Total to reveal all data.

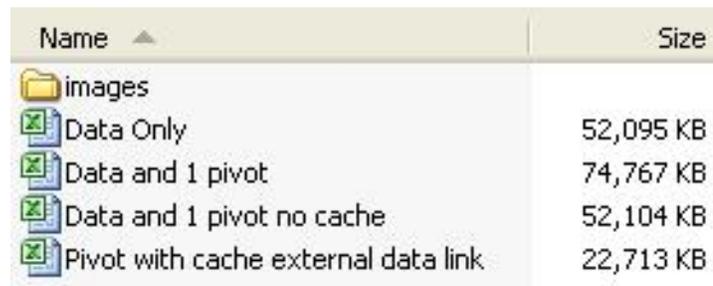
Best of all,

3. Get Data from an External Source

By basing your pivot table on an external source (which may just be another spreadsheet) you have the best of both worlds – the pivot table can be refreshed as the external data table is amended but you also have a reduced file size as only the pivot cache is stored in memory.

Conclusion and file sizes

The image to the right shows the impact of these various options on the file size. As you can see the original data source is a whopping 52MB. This is pure data, it has minimal formatting and no calculations.



Name	Size
images	
Data Only	52,095 KB
Data and 1 pivot	74,767 KB
Data and 1 pivot no cache	52,104 KB
Pivot with cache external data link	22,713 KB

Saving a pivot with its associated cache increases this file size to almost 75MB. (Interestingly it hasn't doubled, which I assume relates to the cache being pure data without formatting but i could be wrong).

By removing the check box “Save source data with file” the file goes back down to the original 52MB so the pivot table itself carries very little extra burden.

The best of all options has been achieved by basing the pivot table on an external spreadsheet – in this case the data only spreadsheet, so that only the cache and pivot are stored. In this instance the file has reduced to more than half that of the original data file.

So my recommendation is to save the data sheet as a workbook and then place all associated pivot tables in a separate workbook – you can allow the pivot table to save the cache (the default position) if you want the benefit of rapid refreshing and plan to have multiple pivot tables with calculations and groupings.

For extreme file size reduction you can combine all of the above techniques, basing the pivot table on the 52MB external data source and then removing the pivot cache reduces the file size to a tiny 17KB – notice K and not M.

Name ▲	Size
images	
Data Only	52,095 KB
Data and 1 pivot	74,767 KB
Data and 1 pivot no cache	52,104 KB
Pivot with cache external data link	22,713 KB
Pivot with cache external data link no cache	17 KB