

## EXCEL



By Bill Jelen

# Formula Speed

Excel 2007 offers 17 billion cells on each worksheet. The Windows operating systems (64-bit versions) support double-digit gigabytes of memory. The PowerPivot feature coming in Excel 2010 will allow you to mash up data sets with 100 million rows. It will be possible to build behemoth spreadsheets with more calculations than ever before. If you've been around since the days of Lotus 1-2-3, you might remember recalculation (recalc) times of 40 seconds or more. Computing power has mostly increased to the point where a typical spreadsheet with less than 65,536 rows could easily calculate in a second or less, so you might have forgotten some of the spreadsheet design rules that can have a dramatic impact on formula speed. The examples in this column can make your spreadsheets calculate hundreds of times faster.

## Measuring Formula Speed

Excel 2007's macro language offers a new Range.Calculate method that allows you to calculate a specific section of a spreadsheet. To get the recalc times for this article, I used Excel 2007 in manual

calculation mode and a macro to calculate one range of formulas.

## Two Approaches to Running Totals

Let's say you have a spreadsheet with 75,000 rows of data in cells A2:E75001 and want to add an accumulated total in column F. There are two common approaches. Think about how you would solve this and mentally vote for your favorite before reading the results.

Method 1 is a dual-formula solution. Cell F2 contains =E2. The rest of the cells in column F contain a formula adding an amount from column E to the running total in column F. For example, cell F3 contains =E3+F2, cell F4 contains =E4+F3, and so forth, down to cell F75001, which contains =E75001+F75000. I refer to this as the "Excel Kindergarten" approach. I despise it because it requires two different types of formulas.

Method 2 starts with a formula of =SUM(E\$2:E2) in cell F2. This is a clever approach that uses a single dollar sign in the reference. As you copy that formula down, the first E\$2 remains locked to row 2, but the second E2 is allowed to

remain relative. In row 75001, the formula will be =SUM(E\$2:E75001). Figure 1 shows the last few examples of each method (for comparison sake, the Method 2 formulas are in column G).

Which version do you prefer? I'll be honest...as MrExcel, I appreciate the brilliance of Method 2. It shows someone who understands relative and absolute references, and it allows every cell in the column to have the same formula structure. It gets my vote.

Let's consider formula speed for the two approaches. In Method 1, there are 74,999 cells that have to look at two other cells and one cell that has to look at one other cell. To calculate that range, Excel has to look at 149,999 cells [(74999 x 2) + (1 x 1)]. In Method 2, the formula in F2 looks at one cell. The formula in F11 looks at 10 cells. The formula in F75001 looks at 75,000 cells. Summing (1, 2, ..., 75,000) means that Method 2 has to look at 2,812,500 cells. Which do you think calculates faster? Method 1 calculates in 0.4936 seconds. Method 2 calculates in 90.21275 seconds. So even though Method 2 is more elegant in creation, it actually performs slower.

## VLOOKUP vs. INDEX/MATCH

Now let's say you have 2,000 rows and 12 columns that need to be filled with a VLOOKUP. The lookup table contains 35,000 unsorted product records. Each column represents a different month.

Setting up those 12 VLOOKUPs can be tedious; even with successful use of dollar signs in =VLOOKUP(\$A5,\$AA\$2:\$AM\$35001,2,FALSE), you'll have to edit the formula in each column to change the third argument from 2 to 3 to 4 to 5, and so on. Some people might fill B1:M1 with the numbers 1 through 12 and then refer to those numbers in the VLOOKUP: =VLOOKUP(\$A5,\$AA\$2:\$AM\$35001,B\$1,FALSE). Others might use the clever approach of =VLOOKUP(\$A5,\$AA\$2:\$AM\$35001,COLUMN(B1),FALSE).

Further, if you're worried about missing values, you might set up an IF statement to check for errors: =IF(ISERROR(VLOOKUP(\$A5,\$AA\$2:\$AM\$35001,2,FALSE)),0,VLOOKUP(\$A5,\$AA\$2:\$AM\$35001,2,FALSE)). This method is effective, but it requires every cell in the report to calculate the VLOOKUP twice. If you have Excel 2007 and can use the

### Table 1. Formula Calculation Times

Method	Calculation Time (in seconds)
VLOOKUP with 2 hard-coded as third argument	43.29
VLOOKUP pointing to B\$1 as third argument	44.31
VLOOKUP with COLUMN(B1) as third argument	43.98
VLOOKUP with =IF(ISNA	94.37
VLOOKUP with =IFERROR(	45.75
INDEX and MATCH	3.68
INDEX and MATCH, Sorted Table	0.03497

new =IFERROR(VLOOKUP(),0), it will only calculate the VLOOKUP once.

A different approach is to insert a new column called "Where" in column B. This column will find the relative record number of each product with =MATCH(A5,\$AA2:\$AM35001,0). From a recal point of view, a column of MATCH functions takes about as much time as a single column of VLOOKUP functions. But you can then fill the 12 columns that used to contain VLOOKUP functions with INDEX functions instead. The formula for month 1 in C2 would be =INDEX(AB\$2:AB\$35001,\$B5). Note that this formula points to a single column of the lookup table. Dollar signs lock the rows to a specific range, but the column is left as relative. This formula can easily be copied to the 12 columns of the report. From a formula

speed point of view, INDEX is lightning fast compared to VLOOKUP. Even with an extra 2,000 formulas, this new range calculates faster

than the fastest VLOOKUP solution.

The time of 3.68 seconds represents a 92% time savings over the normal VLOOKUP. I wouldn't go beyond that result. The one possible further improvement would be to sort the lookup table and use TRUE as the last argument in the VLOOKUP or MATCH formula. While this method is 100 times faster, you run the risk of returning incorrect results if a product is missing from the lookup table.

Over the years, I have used every method in Table 1, often thinking that they were all basically equivalent. When you consider formula speed, the differences are dramatic. **SF**

*Bill Jelen is the author of Excel Gurus Gone Wild and 24 other books about Excel. Send questions for future articles to [IMA@MrExcel.com](mailto:IMA@MrExcel.com).*