VBA Project

Wyatt Nielsen

4/11/2012

Executive Summary

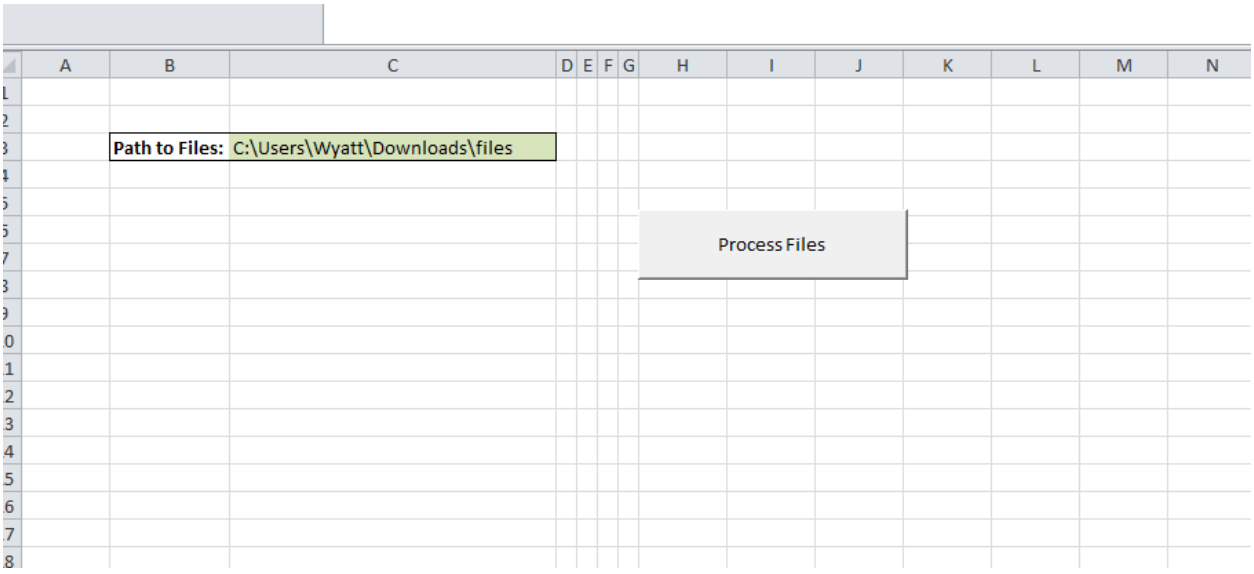
The organization that I work for has several large databases that recently underwent a change that needs to be closely monitored. These statistics are included in a daily html report. To monitor the effect that these changes had on the databases we need to open each html file and look at 2 specific statistics. To really identify any effect of the changes I would need to be able to graph the changes.

To solve this problem I designed this project that uses VBA to cycle through all files in a directory and search each file for the correct statistics, adding them to an excel spreadsheet and graphing the results. To accomplish this I had to use different references(one for the file system object and one for regular expressions)

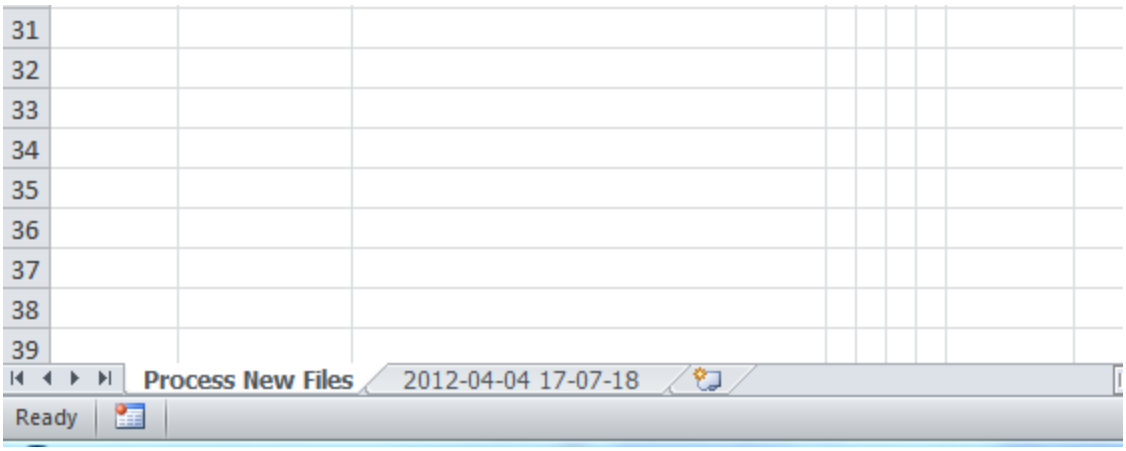
I work with several different databases that manage a large number of users and millions of rows of data. These databases are used in production and so are vital to the organization. Usage and performance statistics are constantly gathered on the databases and nightly the stats from the day are written to an HTML document. Recently there have been some changes within the organization that may have had an effect on performance. To understand what kind of effect the changes had some key statistics needed to be gleaned from several months' worth of files and put into an excel spreadsheet so that they could be graphed and a visual representation of the effect could be shown. This project was to create a VBA solution to this problem. My VBA project will loop through all documents in a folder and search each one for the desired statistics (sequential Read and scattered Read), plot them on a new worksheet, and graph the results.

In implementing this project I ran into several different problems that took some time to overcome. For starters I was unable to search a folder to loop through the contained documents. After spending some time searching I discovered that I had not included the proper extension. After that I was able to quickly loop through all of the files in a folder. Getting the regular expressions exactly right was another challenge that had to be overcome but in the end the project works exactly as designed.

To run this project you simply put the folder containing the documents to parse in the designated cell as shown below:



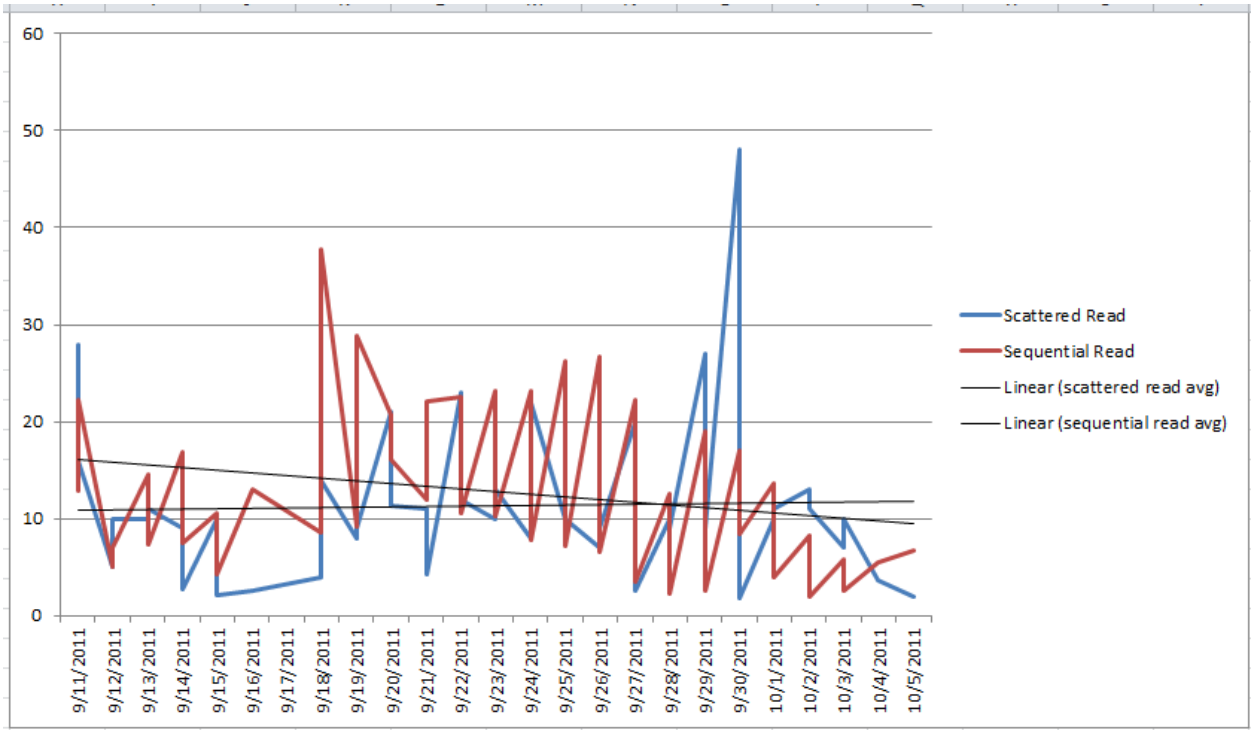
When the process files button is pressed a new worksheet is created with the named by the current time stamp as shown:



This new spreadsheet is where all of the statistics will be plotted and will contain the graphs. Then using a FileSystemObject (included by using a reference to the Microsoft Office Object Library) a list of all file in the folder as written in cell C3. Looping through the files each one is searched using two different regular expressions to extract the two different statistics needed and the statistics are plotted on the new worksheet. Because there could possibly be thousands of files in a run the page updates are shown so that the user knows that it is working. The date is parsed from the filename because each file is named “awrrpt_HOSTNAME_DBNODE_DATE_TIME” eg “awrrpt_Host1_AAAA_20110911-071001”. The data is plotted as shown:

	A	B	C
1	Date	Scattered Read	Sequential Read
2	9/11/2011	28	12.9
3	9/12/2011	5	5
4	9/13/2011	10	14.5
5	9/14/2011	9	16.9
6	9/15/2011	10	10.5
7	9/18/2011	4	8.6
8	9/19/2011	8	9.2
9	9/20/2011	21	20.7
10	9/21/2011	11	12
11	9/22/2011	23	22.6
12	9/23/2011	10	23.2
13	9/24/2011	8	23.2
14	9/25/2011	10	26.3
15	9/26/2011	7	26.7
16	9/27/2011	20	22.2
17	9/28/2011	10	12.5
18	9/29/2011	27	19
19	9/30/2011	48	17
20	10/1/2011	10	13.7
21	10/2/2011	13	8.2
22	10/3/2011	7	5.8
23	9/11/2011	16	22.2

After all of the files have been parsed another subroutine is called that builds a graph that includes both of the statistics and a trendline for each, shown:


60
50
40
30
20
10
0
9/11/2011
9/12/2011
9/13/2011
9/14/2011
9/15/2011
9/16/2011
9/17/2011
9/18/2011
9/19/2011
9/20/2011
9/21/2011
9/22/2011
9/23/2011
9/24/2011
9/25/2011
9/26/2011
9/27/2011
9/28/2011
9/29/2011
9/30/2011
10/1/2011
10/2/2011
10/3/2011
10/4/2011
10/5/2011
Scattered Read
Sequential Read
Linear (scattered read avg)
Linear (sequential read avg)