

## **PBS Tips and Tricks #24 – Purging – how scary should it be?**

PSI's support department is sometimes tasked with helping a client to do an upgrade and we find, not infrequently, a hundred thousand and more vendor history records that must be converted, i.e., read and rewritten, because that file has infrequently been purged or has never been purged at all. This can dramatically increase conversion and update times. For many people this issue is somewhat tinged with worry – what if I need that data? This, of course is a legitimate question. And to some, the safe answer is to never purge data.

Computer speeds and disk sizes have more or less kept up to the demands so that, any overhead resulting from not purging may be tolerable or even acceptable. But is it really to your advantage? For example, after running PBS for a few years you can now find yourself with say, 8000 print-to-disk reports - an unmanageable number of reports. Or, you may have 6 or 8 years-worth of data in your AP or AR Open Item file – and that will increase the run times of the Aging Report and other reports.

There should be a reasonable answer to that question – “when should I purge?” and almost all the time the answer is not “never”.

Answering that question should include the following considerations:

- What, if any, are the statutory requirements?
- What are the business demands that dictate keeping the data?
- What other practical concerns are there?

Another point that will affect the answer is whether or not you have accessible copies of the older data outside of the main accounting system and kept for extended periods. We can do this, for example, by making an “archive company” where we have all the data through say, Dec 31<sup>st</sup> of the last year. Having an archive allows you to re-calibrate your purging considerations to purge more aggressively. You might keep only two years of certain data instead of keeping four. You can do this comfortably because, in the worst case if the data is not in the production system, it IS in one of the archive system(s). And that is not much a worst case. (More on setting up an archive in a later installment.)

Addressing those three points above:

*Statutory requirements.* This is something to discuss with you accountant. Typically, though, it is the Payroll and General Ledger data that will have to be kept for some number of years to allow for audits and other queries from accountants and governmental agencies. Assuming, seven years is safe, having archive companies for past years would allow, for example, keeping only two or three years in your live company because typically, that is the time frame from which you will be frequently using the data and the rest can be in an archive.

*Business requirements.* Every business has its unique requirements and while most businesses will be happy to have, say, two or three years of vendor history and to archive the rest, your business may need to keep records for a longer time. Keep in mind that even if you are not periodically creating archive companies, you may be archiving already in the sense that every year as part of the year-end process you take all the paper work, i.e., for AP and file the vendor invoices, copies of checks etc. to make an archive.

With AP as our example, another way to think about the issue of how much “live” data to keep is to imagine the proportion of lookups and queries you do to your vendor history to check something out. Most of these actions will be current, typically the last 3 to 6 months, but a few may go back a year, etc. Ask yourself - in an average month how often would I have to go to my electronic archive or to my paper archive for a 2-year-old document, or 4- or 5-year-old document? Businesses will differ legitimately, but somewhere in that range there is likely to be a sweet spot, that balances essentially instant access from the production copy of your data, against having to, say once or twice a month, open up the archive company or go to the paper files to check something out. And you don’t have to go “all-out” the first time around. You could be somewhat conservative on the first round of purging and e.g. choose 4 years back and a year later realize the you so rarely had to go to the archive, you could purge to 2 or 3 years. This process of consideration can be applied to any of the files that should be purged and so the thought process crosses all the application modules.

The above paragraph was, in effect, about estimating the downside of the effects of purging/losing instant access to data. The upside to purging the files down will be, for example, cutting the file sizes of the vendor history in half and therefore cutting in half the access time for certain lookups or the run time for reports.

*Practical Concerns.* The positive results of purging are easier navigation through the masses of data you are accumulating - easier because you don’t have to scroll over the old data or have the program read through ten thousand records when what you need corresponds to maybe only 500.

A technical note: many of the PBS reports and queries allow you to select a date range to narrow down or filter out the data to get what you want. This will ignore the old stuff BUT in many cases the program still has to read all the old records to skip over them and ignore them and that simply costs time. Some programs are keyed to date so they can go directly to data and that makes them much less sensitive to these issues. But all files at some point will have to be read sequentially, i.e., from one point to another and this uses time and disk space.)

Since there a lot of people out there who converted to PBS from RealWorld software, there are also a lot of people who brought over everything from that system and then continued to add to it by using PBS. As a result, it is not uncommon to find many years’ worth of data buried in

the files of these systems and cases where the file sizes can be reduced by 80 percent or more by purging. Purging the old unnecessary data can have dramatic results: a report in the AP or AR Aging Report that took a minute or two to run, now gives results essentially instantaneously or in a second or two.

Consider how a minute or two of waiting for a report to finish accumulates and, when that happens a half-dozen times a day, it becomes several days of lost time over a year.

One last aspect of practical concerns. There are numerous instances throughout the PBS system where, in the process of moving the data from one stage to another, the original data is kept. This is not a flaw. It is designed-in because for a short period of time it is useful to have this “duplication” of data. However, in the longer term, this duplicated data is just dead weight and can/should be purged. Two examples should clarify this:

- a) There are five financial modules that regularly create GL distributions in PBS (AR-receivables, AP-payables, IC-inventory, CR-check reconciliation, PR-payroll; in manufacturing the inventory IM module does something similar to IC). Each module has its own storage of the GL distribution data that it is responsible for. If you don't have General Ledger, these sets of distribution transactions must be accessed through that module and these distribution files are the only source of this data.

If you do have General Ledger and most of our customers do, these files should be considered as just temporary holding areas until the data is moved into the General Ledger. (Those of you who are doing GL will be familiar with the monthly task of interfacing the data from the above sub-ledgers into the GL.) This means that once the data is interfaced to GL and a little time has passed after closing the month, there is no reason to keep these distributions and they can be purged.

- b) In Accounts Payable, PBS keeps track of all payments against vouchers in the Open Item file. It also keeps this information in the Vendor History file. This duplication is done for a reason. When you void a check, if the original vouchers for the items on that check are in the Open Item file, then the program can simply re-instate that voucher as “open” (as well as doing all the other things involved with voiding the check). This means there is an advantage to keeping the fully paid stuff around for some time.

However, as a convenience for users so they just see the open items that are up for payment in the Aging Report, the program will suppress the printing of any items that are fully paid. This keeps the AP Aging report “clean”. But the downside to this is that the fully paid vouchers are “hidden” so their presence after payment is less obvious.

In this example, the duplication of the data is useful for a period of time. What is that time?

Most banks will stale date a check older than some number of months. Whatever that number is, after that period, having these fully paid vouchers kept on file makes no sense and there is no down-side to removing them.

So, purging the AP Open file of fully paid items should be done for entries that are something like 6 or 12 months old.

(Voiding a check where the relevant Open Item entries have been purged is still possible and will be dealt with in a later installment.) We will also walk through some of the details about purging the various files in PBS.