THE 3000 NEWS Wire
Independent Information to Maximize Your HP 3000

Latest Adager: Mastering Masters

Version tames MDX complexities, provides lightning-fast master dataset re-hashing

Review by John Burke

Adager. Everyone knows about Adager. Just about anything you want to do to your database structurally you can do with Adager — from the company called Adager. Quickly, easily and correctly. And their support is second to none. I've personally written at least two reviews of Adager over the years. So why another? Two reasons. Version 000121, which was used for this review, includes:

• support for the latest IMAGE enhancement, automatic dynamic master dataset expansion (MDX); and

• an enhanced native-mode rehashing module for master capacity changes interfaced with MDX.

Since the advent of in-place capacity changes for detail datasets (DDX), master dataset capacity changes are the one periodic database maintenance operation that requires significant downtime. In order to avoid running into unscheduled downtime because of a master dataset getting suddenly full, users are now typically significantly over-sizing master datasets. This of course wastes disk space.

But more importantly, while hashed access doesn't suffer, serial access to these datasets definitely does because of all the "empty" entries. Use of MDX, coupled with Adager's new super-fast rehashing module, will empower users to better and more efficiently manage their disk resources, processing times, and downtime. In other words, users will be able to better manage their databases.

Master dataset rehashing

Let's look at rehashing first, because it is of wider immediate use and interest and is conceptually much easier to deal with than MDX. When you change the capacity of a master dataset, whether manual or automatic, all the entries must be "rehashed". This is similar to reloading the entire dataset from scratch. If there are millions of entries, this can clearly be a very time consuming-process.

Simply put, Adager has a new native-mode rehashing engine for master dataset capacity changes that is revolutionary in its performance gains. At several customer locations where it has been tested, Adager reports the new module has shown performance improvements of up to 25 times over the previous rehashing engine. The new master dataset capacity change module works with any version of IMAGE and can be of immediate benefit to all TurboIMAGE users. So what more is there to say? Let's test it.

Lab Results

Figure 1 shows my test configuration and my test results. They are spectacular in both CPU and wall time (particularly the very important wall time, which corresponds to database downtime).

Going from over 8 hours down to about 1.5 hours for dataset 2 blew me away. Bottom line, I saw four to five times improvement in wall time. I ran this test to show the relative improvement of the new hashing module over the old, keeping both system and database constant. However, this is not a true real-world situation. This is an old system limited by a slow CPU, small memory and several HP-IB disc drives. The advantage of using such a minimal system, however, is that I can safely say something like "your mileage may very — but you are almost certain to see even better performance than I achieved on the test machine."

Case in point, the following are statistics from a real-world application of the new rehashing module. Unfortunately, because it was a production system, we do not have any "before" numbers to compare.

System: HP 3000/959-400 with 2 Gb memory
OS: MPE/iX 5.5 PP6
Dataset: SERVICE-A (A)
Entries: 6,792,899
Old Cap: 8,239,349
New Cap: 11,000,027
Time Elapsed: 38 minutes
CPU: 1865 seconds

Adager performed at a rate of close to 11 million entries per hour! This is spectacular. SERVICE-A is part of the Amisys HEALTH database. The data item type for the search field is X16. In my "lab" test, the hashing rates were 479,000 entries per hour for dataset 1, and 2.1 million entries per hour for dataset 2. These statistics illustrate how difficult it is to project a time estimate for a master capacity change, especially with the wide range of HP 3000 CPUs out in the field. They also validate my earlier statement that "your mileage may very, but you are almost certain to see even better performance than was achieved on the test machine."

The tested release of Adager also includes a feature that will allow users to obtain a snapshot of the vital statistics as a by-product of the rehashing process. Adager will save the statistics in a file called "savestat" in the database's group and account. Future versions of Adager will also include a "data layout statistics" section in the report that will itemize such things as percent primaries, longest run of primaries, longest run of entries, and longest synonym chain.

MDX

From the MPE/iX 6.0 Communicator: "The MDX feature allows dynamic expansion of a non-jumbo dataset during DBPUT when the data set has approximated its current capacity and DBPUT would fail unless the dataset is expanded. As in DDX, the capacity parameters, which are maximum capacity, initial capacity, and increment (optional), used for dynamic expansion, must be set prior to the actual expansion. For new databases, these parameters can be specified in the CAPACITY statement of the schema definition to be processed by DBSCHEMA. For existing databases, third-party tools that support MDX need to be employed."

MDX was first made available to IMAGE

Figure 1

System: 3000/948 with 96MB memory
OS: MPE/iX 6.0 with PowerPatch 1
IMAGE/SQL: C.07.21
Database: Superdex-enabled with 69 sets including the Superdex Index (SI) set. This is a copy of a production database from a 959/400.

	Dataset 1	Dataset 2
Name	PART	AP-INVOICE-MSTR
Set #	16	32
Type	M	M
Initial Capacity	375,787	5,500,003
# Entries	255,539	3,467,340
% of max	68	63
# of paths	11	0
# Fields	63	2
Key	X20	X26
TPI indexes	1	0
Blocking factor	1	40
Length	392	763
Sectors	1,503,152	825,008
MB	384	211
New capacity (50%)	511,087	6,934,687
New sectors	2,044,352	1,040224
New MB	512	256
"old" version		
CPU seconds	1,420	9,878
Wall time (min)	112	502
000121 version		
CPU seconds	909	4,921
% of "old" version	64	50
Wall time (min)	32	98
% of "old" version	29	20

users as of IMAGE version C.07.01. In its initial release, HP's implementation required that the hashing capacity be a multiple of the blocking factor. In almost all cases this forced users wishing to enable MDX on existing datasets to do a rehashing of the existing entries, a process that could take several hours or days. HP later enhanced its implementation (in large measure because of prompting from Adager) to remove this restriction. This revision has been available since version C.07.18 of IMAGE. Note that MDX is not currently supported on jumbo datasets (nor is DDX for that matter).

Fred White of Adager (and for anyone who does not already know, a member of the original IMAGE development team) wrote a detailed article discussing both MDX and DDX in the NewsWire's December 1999 issue. I highly recommend it for anyone who is responsible for maintaining TurboIMAGE databases.

Adager supports MDX on any version of MPE that supports IMAGE versions C.07.18 or newer, though it is highly recommended that you be on at least C.07.23 before using MDX in production because of MDX problems in earlier versions of TurboIMAGE. Unfortunately, this means you will likely have to apply a patch to your system in order to use MDX since MPE/iX 5.5 PP7 and MPE/iX 6.0 only come with version C.07.14 of TurboIMAGE. And even MPE/iX 6.0 PP1 only has version C.07.18 of TurboIMAGE.

Adager allows you to enable MDX on master datasets that do not have it, change any of the existing MDX parameters, and remove MDX on master datasets currently enabled. During any MDX-related work, Adager analyzes your MDX operation and will minimize the amount of work that your request requires. Adager supports MDX on all relevant functions. This means that you can do structural changes to a master dataset with MDX and preserve its MDX properties. If necessary, Adager will adjust some of the MDX parameters to maintain consistency in their values. You can also perform any Adager maintenance, diagnostics, therapy or browsing functions on master datasets enabled for MDX without having to request any special instruction. And if you have B-trees attached to an MDX master, Adager will automatically re-index them for you when necessary.

Two examples where you can save hours of downtime with Adager and MDX:

• You have a master dataset with millions of entries that you wish to enable with MDX. If you specify the initial capacity (also known as the hashing capacity) equal to the current capacity, Adager will recognize those cases such as this, where there is no need to rehash the dataset.

• You have a dataset already enabled for MDX and wish to increase its maximum capacity or change the size of the increment. Adager does it in seconds.

In its support for MDX, Adager automatically enforces a rule it developed for DDX, to make sure that your datasets will always have a maximum capacity allocated to allow a complete increment, even if it is the last one before the dataset is full.

I posed the following question to Adager principals Alfredo Rego and Rene Woc: "Very few users have attempted to deploy MDX. How is your customer base reacting to the introduction of MDX?"

Rego said, "MDX has not quite taken off yet, due to initial problems having to do with the original requirement for an initial hashing capacity that had to be a multiple of the blocking factor. Even after HP corrected this design oversight, there were other problems. Things seem to have settled down now, but it takes a while between the time that the IMAGE Labs release something and the time that production users actually incorporate it. MDX is still in this transition phase."

Woc said, "I think MDX will need some time for the user to understand how it works. It will definitely help users who only have a few time windows for maintenance during the year. That was the main reason for requesting that one could enable MDX without the need for rehashing. MDX also has the potential of helping, performance-wise, by practically eliminating the possibility of having entry clusters that in the past have caused very severe performance degradations.

He added, "As MDX moves into general use, HP might also provide the means for a user-controllable maximum cluster length setting. This is a hard-wired parameter right now, but could be useful in performance tuning a database. So far, HP has not committed to supporting DDX and MDX on datasets greater than 4 gigabytes. Neither MDX nor DDX is currently supported with jumbo datasets. As 'large files' become available in MPE/iX 6.5, I'm sure HP will feel the pressure to support

3000 Test Drive Road Report

Adager
Version 000121

Adager Corporation
The Adager Way
Sun Valley, Idaho 83353-3000
Phone: 208.726.9100
Fax: 208.726.8191
E-mail: info@adager.com
Web: www.adager.com

The latest version of Adager, the Adapter/Manager for IMAGE/SQL databases, adds complete support for automatic master dataset expansion (MDX) and features an enhanced, lightning-fast master dataset rehashing (capacity change) module.

Adager runs on all versions of MPE. In order to take full (safe) advantage of MDX, you should be running version C.07.23 of TurboIMAGE or later. Adager Model 1 (for day-to-day maintenance functions such as capacity changes and dataset repacking) costs $2,300 with annual maintenance after the first year of $300. Adager Model 2 (which includes everything in Model 1, structural change modules, diagnostic and repair functions) costs $7,500 with annual maintenance after the first year of $1,000. Discounts may apply for organizations with multiple systems.

'single large files' as dataset files and support for dynamic expansion will follow naturally. MDX is not a simple concept or implementation but it has a very nice potential to help IMAGE users.

"Being able to turn on (enable) MDX on an existing master without having to rehash the existing entries is a feature that the original implementation of MDX did not allow. According to the TurboIMAGE manual (which describes the original implementation), the initial capacity is adjusted to represent an even multiple of the blocking factor. In MDX, the initial capacity of a master is the 'hashing capacity.' In practice there are no master datasets that have their capacity as a multiple of the blocking factor, except when the blocking factor is 1. In the original implementation, this meant that to enable MDX on a multi-million entry dataset one would have to rehash the entries.

"As of TurboIMAGE C.07.18, however, the initial (hashing) capacity was allowed to be any number. This enhancement allowed us to enable MDX on an existing dataset without having to rehash the entries (by specifying the initial capacity to be equal to the pre-MDX capacity). If a set does not have MDX and you are enabling MDX, Adager will not rehash if the MDX initial (hashing) capacity is equal to the pre-MDX capacity. If a set has MDX, Adager will not rehash whenever the initial capacity does not change. In either case, Adager will adjust the Maximum Capacity to be a multiple of the difference between the [(Maximum Capacity - Current Capacity) / Increment], after rounding the requested increment to be a multiple of the blocking factor."

So what happened in my test examples? If I chose to expand capacity to the same values as before, but used Adager to enable MDX, the "expansion" was performed in seconds. All you need do is specify initial capacity = current capacity in the dialog.

As always, Adager will protect you from the various idiosyncrasies of different versions of TurboIMAGE. Versions of TurboIMAGE older than C.07.18, for example, will trigger specific Adager messages if you attempt to assign a hashing capacity (also known as the "initial capacity") which is not a multiple of the master dataset's blocking factor.

Final word — and a little truth in reporting

Okay I'll admit it, I'm biased. I've used Adager, the product, for at least 20 years at several different job stops — I'll never forget explaining to the executive committee back in 1980 why I absolutely had to have a software product from a company in Guatemala.

"I've used Adager for at least 20 years" actually says it all. I had choices, but I stayed with Adager. No product is better. But more importantly, Adager — the company and people — have set an example of customer and technical support that is absolutely unrivaled. While I do not wish problems on anyone, it is only when you do have that problem at 5 AM on a Sunday that you discover how good the support from Adager really is. And can appreciate the unique relationship Adager has with its customers.

This version of Adager continues the Adager tradition: improving performance and providing support for TurboIMAGE's continuous enhancements.

John Burke, who edits the NewsWire's Hidden Value and net.digest columns, is a member of the MIS staff at Pacific Coast Building products with more than 21 years of HP 3000 experience.