

Checklist to get the most out of x/EFS

Hanns-Joachim Uhl

The following table provides a checklist that can be used to help tune xSeries Enabled for System/390 (x/EFS) systems for best performance. IBM's x/EFS systems were formerly known as *NUMA-Q Enabled for S/390* or *NUMA-Q EFS* systems. The points listed in the table are explained in notes following the table. For more information about x/EFS systems, see the red-book *NUMA-Q Enabled for S/390: Technical Introduction*, SG24-6215, dated December 2000.

Resource	How to check?	Remarks
All Intel processors online to ptx?	ONLINE command in ptx	All processors should be available. (Note 1)
Any paging by ptx?	MONITOR command in ptx	No paging should occur (Note 2)
CACHESIZE in FLEX-ES	Check FLEX-ES configuration file	Default 1024. (Note 3)
Cache miss rate in FLEX-ES	D CACHESTATS command	Less than 4%. (Note 4)
TRACKCACHESIZE in FLEX-ES	Check FLEX-ES configuration file	(Note 5)
Disk cache for each CU in FLEX-ES	D CKDCACHESTATS command and configuration file	(Note 6)
Writeback or writethrough for FLEX-ES disks?	D DEVSTATE and configuration file	Default is writeback. (Note 7)
Dedicated Intel processors for FLEX-ES?	Check FLEX-ES configuration file	Default: not dedicated (Note 8)
VSE/ESA: DASD Fast Write and NVS	CACHE UNIT=cuu, STATUS command	No effect (Note 9)
VM/ESA: DASD Fast Write and NVS	Q DASD DETAILS cuu command	No effect (Note 9)
OS/390: DASD Fast Write and NVS	DS P,cuu command	No effect (Note 9)

Note 1. In general, all Intel processors in the xSeries EFS system should be online and this is the default. This should not be changed. Processor status can be verified with the ONLINE command in ptx. For example, in a two quad system, the command and output might be:

```
# online
online: 0-5
offline: 6-7
```

In this example two Intel processors are offline (numbers 6 and 7) and all other Intel processors are online. The command `online 6`, for example, would bring processor 6 back to the online state.

Note 2. In general, no paging should occur in ptx because the S/390 CPU emulation process used by FLEX-ES is very sensitive to any ptx paging. You can check whether paging occurs in ptx by using the `monitor -f -i 10` command. In this command ‘f’ indicates that two monitor screens are produced and the ‘f’ key (on the keyboard) is used to toggle between the screens. The ‘-i 10’ parameter means that the monitor information will be updated every 10 seconds; the default is one second. Any ‘f’ keystrokes will be accepted only at the update intervals (10 seconds in this example). Examples of the screens are:

(First Screen)

```
P# +-----+-----+-----+-----+-----+-----+-----+-----+-----+
0 |xxxxx
1 |xxxxxxxx
2 |
3 |
4 |
5 |
6 |
7 |
P# +-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

(Second screen)

```
Total User Time 0      Free Page Recs 0      Message Ops 0
Total System Time 0    Dirty Page Recs 0    Semaphore Ops 0
Total Time 0           Page Ins 0           Quad[0] freemem 1217976
Number of Procs 255    Pages paged in 0     Quad[1] freemem 1769056
Swapped Procs 0       Page Outs 0
Running LWPs 0        Pages Paged Out 0
Runnable LWPs 1       Swap Ins 0
Fast Wait 0           Pages Swapped In 0
Sleeping LWPs 254     Swap Outs 0
Swapped LWPs 0        Pages Swapped Out 0
LWPs FS IO Wait 0     FS Blk Reads 0
LWPs Phy IO Wait 0    FS Blk Writes 0
    (and so forth, with more system statistics)
```

This example indicates that no ptx paging is occurring. If paging is occurring, you can install more system memory, reduce the emulated S/390 memory sizes, and/or ensure that unnecessary ptx processes are not running.

If the `monitor` command does not display output correctly, you may need to set the terminal type. The command `export TERM=vu320` in `ptx` should work.

If you are running VM/ESA (under FLEX-ES) and the `monitor` command always indicates 100% CPU busy for the Intel processors, you should consider using the FLEX-ES *FEATURE LPAR* option. With this option, the `monitor` command will more accurately indicate the busy state of the emulated S/390. This is discussed in Chapter 6 of the redbook mentioned earlier.

Note 3. The “`cachesize(nnn)`” parameter defines the size of memory used by FLEX-ES for S/390 instruction caching. To calculate the total xSeries memory used by this cache, use the formula:

$$\text{MBytes of xSeries memory per emulated S/390 processor} = (\text{nnn cache size}) * 11 / 1024$$

For example, if you specify `cachesize=8192` and if you are emulating three S/390 CPUs, the total xSeries memory used for the FLEX-ES processor cache is 264 MB. This is computed as follows:

$$\begin{aligned} \text{Each emulated S/390 CPU uses } & (8192 * 11/1024) = 88 \text{ MB} \\ \text{Three S/390 CPUs use } & 3 * 88 \text{ MB} = 264 \text{ MB} \end{aligned}$$

This is a substantial amount of xSeries memory and must be considered when planning your system.

The default `cachesize` value is 1024, and must be a power of two. That is, valid cache sizes are 1024, 2048, 4096, and so on. If the default size (1024) is used in the above example, the total xSeries memory used would be 33 MB. The default value is considered adequate for most environments. If you experiment with larger values, be certain to monitor the `ptx` paging rate. A large `cachesize` that drives `ptx` into paging will have a negative overall impact on performance.

Note 4. The cache miss rate in the S/390 instruction cache of FLEX-ES should not be higher than about 4%. You can check this with the FLEX-ES command `display cachestats`. An example of this command and results is:

```
flexes> d cachestats
Cache hits (ml): 604325140/1122334222 54%
Cache hits (fb): 11580609/1122334222 1%
Cache hits (fbt): 439740495/1122334222 39%
Cache misses: 66687978/1122334222 6% <=== cache misses data
Cache compiles: 50323445/66687978 75%
Cache bypasses: 33466702/66687978 50%
  (more statistical data follows)
```

The cache miss rate for the S/390 instruction cache can be found in the fourth line. The recommended maximum is 4%. If the displayed value is 4% or less, there is no need to increase the `cachesize` value. If the number is greater than 4% (over several measurements), you should consider increasing the cache size, but not to the extent you cause significant `ptx` paging.

You can clear and reset these statistics with the FLEX-ES command `clear cachestats`.

Note 5. The “`trackcachesize=nnn`” parameter may be used for each FLEX-ES emulated disk control unit to provide more cache for disk I/O. By default, 15 tracks (emulated ckd tracks) of cache space are automatically reserved and dedicated for every 3380, 3390, or 9345 disk drive emulated by FLEX-ES. (Other values apply for other device types.) The `trackcachesize` parameter can allocate more cache, above this default amount for each emulated disk drive.

The computation is slightly indirect. Suppose you specify:

```
options `trackcachesize=2048`
```

for an emulated S/390 disk control unit, and this control unit has 16 x 3390 drives defined. The default number of disk cache tracks (for the 16 disks on the control unit) is 16 x 15 = 240 track caches. You specified 2048 tracks in the options parameter. FLEX-ES will compute

```
2048 - 240 = 1808 additional track caches
```

and uses these additional 1808 tracks of cache space as floating caches used as necessary for any/all drives on the emulated control unit. Each 3390 track cache requires about 58 KB storage (xSeries storage); this example (2048 caches) would use about 118 MB of xSeries storage. A reasonably large disk cache can improve performance, but it should not be increased to the point where it provokes ptx paging.

Note 6. You can check disk cache effectiveness with the FLEX-ES `display ckdcachestats cuu` command, where `cuu` is one `cuu` of the control unit. An example is:

```
flexes> d ckdcachestats 300
ADDRESS  READS  WRITES  CACHE HITS  DEDICATED LINES  LINES USED
0300     52     0      51 (98%)    15                1 (0%)
0301     52     0      51 (98%)    15                1 (0%)
0302     10     0       9 (90%)    15                1 (0%)
0303     10     0       9 (90%)    15                1 (0%)
0304      4     0       3 (75%)    15                1 (0%)
0305      4     0       3 (75%)    15                1 (0%)
      (and so forth)
```

You can clear and reset these statistics with the FLEX-ES command `clear ckdcachestats cuu`, where `cuu` is the appropriate emulated control unit identifier.

Note 7. *Writeback* of data through the FLEX-ES disk cache is the default setting and provides the best performance. However, it can be unsafe from the S/390 viewpoint in the event of a power outage or system failure. We generally recommend using the option “`devopt writethroughcache`” for every emulated disk. This tells FLEX-ES to report a disk write operation is complete (using the normal interrupts associated with S/390 channel operations) only after the data is actually written to disk. Although *writethrough* is slower than *writeback*, it is safer in terms of recoverability of data at the S/390 level.

You may want to consider the need for *writeback* for each emulated disk volume. Some volumes, perhaps solely for temporary work data sets, might be candidates for *writeback* and its better performance. When a disk is managed in *writeback* mode, outstanding data buffers are usually flushed to the disk within 5 to 10 seconds. *Writeback* is the default setting and cannot be explicitly specified in the FLEX-ES configuration file.

You can check the current setting of each emulated disk with the FLEX-ES command `display devstate cuu`, where `cuu` is the emulated address of the disk drive. For example:

```
flexes> d devstate 122
Filename: /dev/vx/rdisk/s390dg/33903V02 State: OPEN, READY
Options: trackcachesize=15
flexes> d devstate 121
Filename: /dev/vx/rdisk/s390dg/33903V01 State: OPEN, READY
Options: trackcachesize=15,writethroughcache
```

In this example, disk 122 is used with the default *writeback* (because it is not shown in the options parameters) and disk 121 is used with *writethrough*. These options cannot be dynamically changed. To change them, you must change the FLEX-ES configuration file, recompile it, and restart the emulated S/390.

Note 8. *Dedicated processors* under FLEX-ES does not have the same meaning as under VM/ESA or in an LPAR environment. A setting of *dedicated* will improve Intel's IA32 cache utilization for the FLEX-ES emulation program by preventing cache contamination by other ptx processes. The dedication is not absolute. Certain ptx system processes can still run on the dedicated IA32 processors, but the more general processes in the ptx system are prevented from running on these processors.

A specification of *dedicated* excludes the possibility of sharing that IA32 processor with another instance of FLEX-ES. For example, if you are licensed for a single FLEX-ES processor and you want to run both a test and a production copy of VSE/ESA (that is, emulate two S/390 systems), you cannot specify *dedicated* for either instance's configuration file.

Fundamental Software recommends that, for any one instance of FLEX-ES, either all the processors be *dedicated* or none of the processors be *dedicated*. Do not have a mixture of dedicated and non-dedicated processors defined for a FLEX-ES instance with more than one S/390 CPU specified.

Note 9. Although DASD Fast Write (DFW) and Non-Volatile Storage (NVS) can be set, by an S/390 operating system, for every disk behind a 3990 control unit, these settings do not have any effect on the actual I/O behavior of disks on an x/EFS system. Although FLEX-ES accepts the CCWs for activation of DFW and NVS and indicates successful activation back to the S/390 operating system, actual I/O operation of emulated disk drives is affected only by the *writethrough* or *writeback* state of the emulated drives. See note 7 for a more complete discussion of this topic.

Trademarks

VSE/ESA and xSeries 430 are trademarks of International Business Machines Corporation in the United States or other countries or both.

IBM, NUMA-Q, S/390, ptx, System/390, OS/390, and VM/ESA are registered trademarks of International Business Machines Corporation in the United States or other countries or both.

FLEX-ES is a trademark of Fundamental Software, Inc.

Intel is a registered trademark of Intel Corporation.

Other company, product, and service names may be trademarks or service names of others.