IBM IT Infrastructure

SFS Performance Management Part 2: Mission Possible

Version 3.0 – see https://www.vm.ibm.com/library/presentations/ for latest version.

Bill Bitner
z/VM Development Lab Client Focus & Care
bitnerb@us.ibm.com

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

Db2*
DirMaint
DS8000*
ECKD
FICON*
FlashCopy*
FlashSystem
GDPS*
ibm.com
IBM Cloud*
IBM eserver
IBM (logo)*
IBM Z*
LinuxONE*
LinuxONE Emperor
LinuxONE Rockhopper
OMEGAMON*
PR/SM
RACF*
System z10*
System 390*
WebSphere*
XIV*
z13*
z13s
z14
z15
Z10 BC
z10EC
z/Architecture*
zEnterprise*
zPDT
z/OS*
zSecure
zSeries*
z/VM*
z Systems*

* Registered trademarks of IBM Corporation

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
IT Infrastructure Library is a Registered Trade Mark of AXELOS Limited.
ITIL is a Registered Trade Mark of AXELOS Limited.
Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.
Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
UNIX is a registered trademark of The Open Group in the United States and other countries.
VMware, the VMware logo, VMware Cloud Foundation, VMware Cloud Foundation Service, VMware vCenter Server, and VMware vSphere are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.
Other product and service names might be trademarks of IBM or other companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.
This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g, zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the “Authorized Use Table for IBM Machines” provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html (“AUT”). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Acknowledgements

- My thanks to various folks for helping pull this material together.
  - Ed Bendert
  - Melissa Carlson
  - Wes Ernsberger
  - Sue Farrell
  - Jim Sculley
  - Butch Terry

Quick Review

- Good planning is a must
- Preventative Tuning
  - CP tuning considerations
  - CMS tuning considerations
  - Disk placement
  - VM data spaces
  - Proper sizing
- Monitor Performance to establish a baseline
- Use the documentation from z/VM Library
  - CMS File Pool Planning, Administration, and Operation
  - Performance
  - CMS Application Development Guide
  - Performance Toolkit Reference

SFS Strengths

- File-level sharing and security

- Disk space management

- Improved file referencing
  - Hierarchical directories
  - Direct file referencing
  - File aliases

- Distributed (remote) file access

- High-level language callable API – Callable Services Library (CSL)

- Data integrity through workunit concepts

SFS vs. Minidisk

- Processor requirements
  - Increase with SFS in proportion to rate of file operations
  - Typical interactive CMS workload, processor time per command increases ~15%

- Real memory requirements
  - Base per-user increase in memory requirements for SFS, but relatively low in 2020 standards
  - Increase can be minimized or reversed by exploitation of SFS file referencing capabilities and VM data spaces

- I/O requirements
  - Similar, just moves from end users to SFS file pool server.

Estimated Processor Requirements

- Proportional to file I/O rate
  - Minidisk I/O is mostly through Diagnose instructions x’A4’ and x’A8’
  - SFS I/O is counted in the server or in normal monitor Block I/O calls (*BLOCKIO)

- Rough estimate through:

$$\texttt{Processor/Command Increase} = \frac{\texttt{Virtual I/O Rate}}{\texttt{Processor Utilization x MIPS}} \texttt{ X } 6\%$$

- When I/O is moved to dircontrol directory exploiting data spaces, the processor usage increase is close to 0%

- Even if you move all user data to SFS, a lot of the I/O will remain to minidisk for things like CMS 190 mdisk

Estimated Processor Requirements - Example

- Processors rated at 100 MIPS
  - From your favorite cheat sheet – or CPUMF from z/VM Download Page to measure actual values
- Current workload does 150 virtual I/Os per Second
  - From Toolkit FCX100 CPU report for system wide number or FCX112 User report for individual virtual machines
- Current workload is 250% busy (each core = 100%, so 2.5 cores)
  - From Toolkit FCX100 CPU report for system wide number or FCX112 User report for individual virtual machines

$$\text{Processor/Command Increase} = \frac{150}{2.5 \text{ x } 100} \text{ X } 6\% \quad = \quad 3.6\%$$

Estimated Memory Requirements

- Very trivial in today’s standards
- Base costs ~ 1800 4KB pages = 7MB
- Per SFS user ~ 4 4KB pages
- On what does it depend?
  - Start-up parameters and other tuning
  - Number of concurrent active workunits
  - Use of two-phase commit
  - Use of data spaces
  - Number of files accessed as filemodes at any time

$$\text{Additional Memory (MB)} = \frac{1800 + (\text{Users} \times 4)}{256}$$

Access Performance – File Control Directory

- First access by any user in the file pool – slower because the file pool server has not cached any of the needed information yet

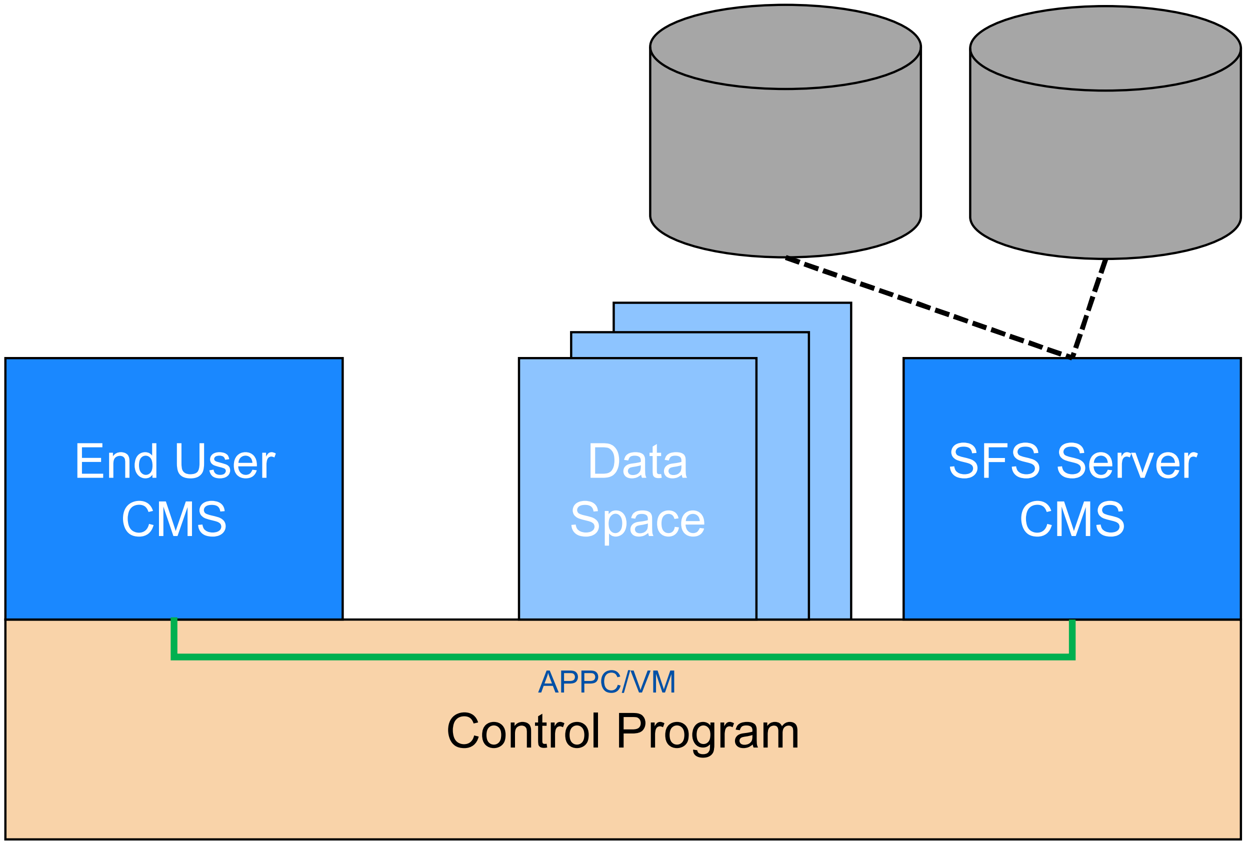
- First access by each user – slower for that user as building structures for all files on the directory requires getting all the information from the server.

- Proportional to the number of files on directory.

- Proportional to the number of files in the directory that require authorization checking
  - Ones you do not own or are not public

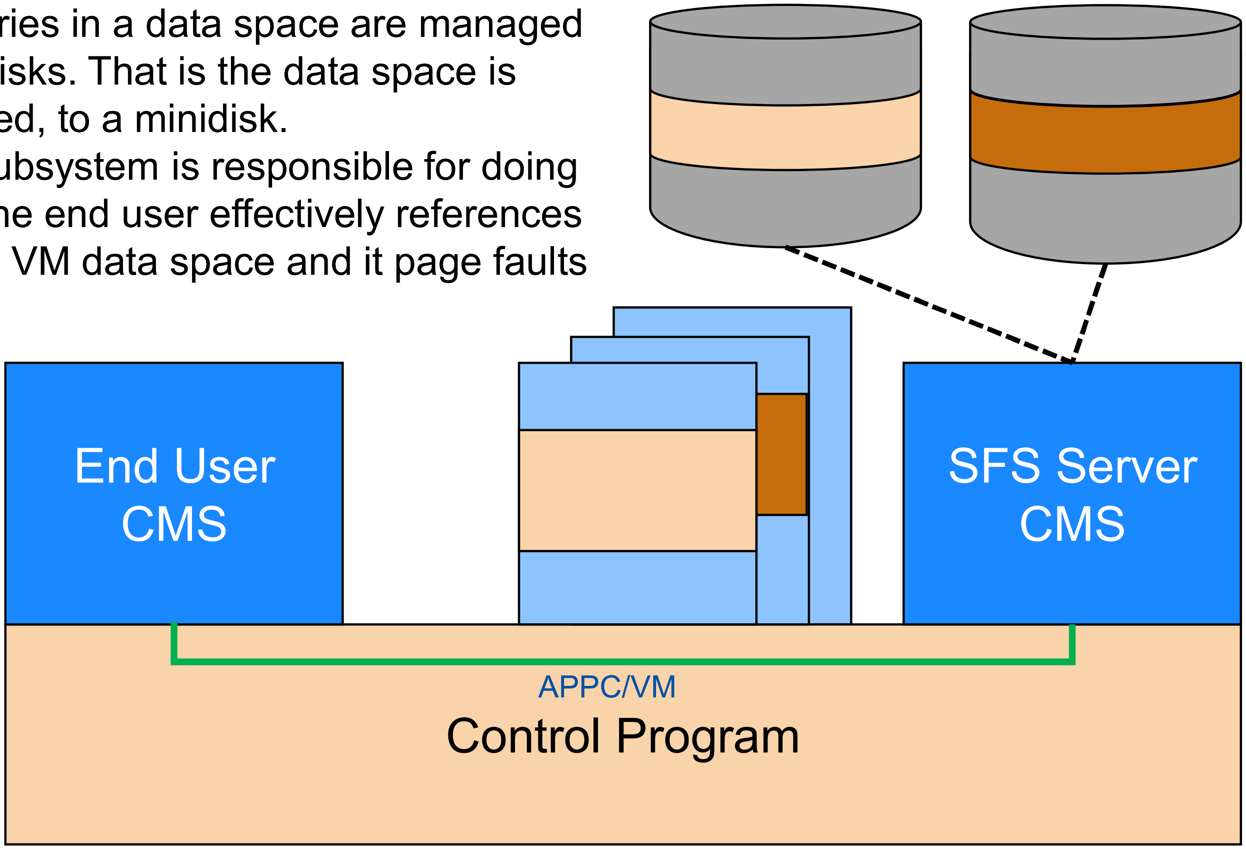
- Directory information in end-user updated when required so that changes are reflected.

VM Data Spaces



VM Data Spaces

- Dircontrol directories in a data space are managed as mapped minidisks. That is the data space is aligned, or mapped, to a minidisk.
- The CP paging subsystem is responsible for doing the I/O. CMS in the end user effectively references an address in the VM data space and it page faults in if not resident.



VM Data Spaces

- Usage considerations
  - Most benefit from highly used shared R/O or read-mostly data
  - Group updates to minimize multiple versions
  - End users should run in XC[1] mode virtual machines for most benefit
  - Consider using different file pools for R/O vs heavy R/W activity

- Performance advantages
  - Relative to minidisk file system
    - Performance similar to minidisk with minidisk cache
  - Relative to SFS without data spaces
    - End user retrieves data from shared virtual memory
    - Most communication overhead with SFS server eliminated
    - End users get data directly from data spaces
    - Control blocks describing files (FSTs) are shared in the data space

[1] With service to z/VM 7.2, z/CMS can run in an XC mode virtual machine. See VM66201 for details.

Data Space Usage

- A separate file pool server for data space usage has advantages

- File pool server should have very little activity
  - Activity is required for initial handshaking
  - A lot of activity is a sign that something is wrong

- Things to watch
  - Users not accessing as R/O
  - Number of data spaces available exhausted
    - CP directory XCONFIG statement determines amount of storage and number of data spaces
    - Monitor data or CP IND USER server_name EXP gives number of data spaces
    - QUERY ACCESSORS *filepool* (DATASPACE command
    - Check for multiple copies of directories
  - Memory usage can be determined by performance data or CP commands
    - INDICATE USER *userid* EXP
    - INDICATE SPACE USER *userid*

Access and Memory

Style	Memory Impact	Constraint below 16MB
Minidisk without SAVEFD	Yes	Yes
SFS filecontrol	Yes	Yes
Minidisk with SAVEFD	No	Yes
SFS data space	No	No

- Control blocks describing memory (FSTs) are slightly larger for SFS than minidisk

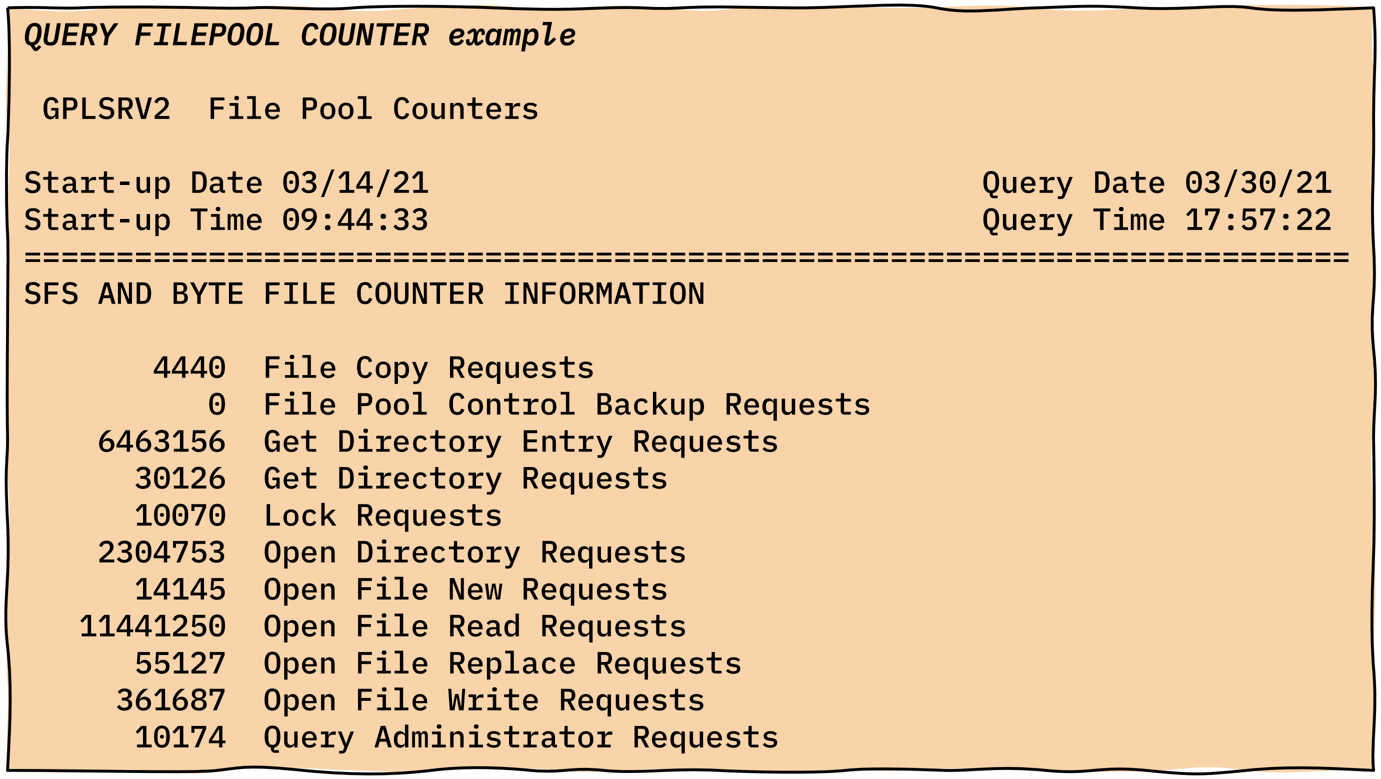
- Benefits of managing SFS dircontrol directories over SAVEFD (places file directory information for a CMS formatted minidisk into a saved segment).

- SFS FSTs in a data space must still be below the 16MB line, but are not in the base address space

- SAVEFD savings example
  - 2000 files for 1000 users on minidisk, saves ~122MB of real memory

SFS and CRR Counters

- Available through Monitor or QUERY FILEPOOL commands
  - REPORT
  - AGENT
  - COUNTER
  - LOG
  - MINIDISK
  - OVERVIEW
  - STORGRP
  - …
- Snapshot of about 200 running counters
- Performance Toolkit reports on counters

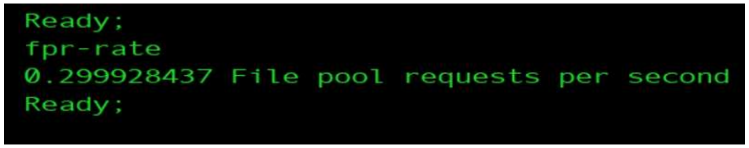


QUERY FILEPOOL COUNTER example

```
 GPLSRV2  File Pool Counters

Start-up Date 03/14/21                                     Query Date 03/30/21
Start-up Time 09:44:33                                     Query Time 17:57:22
==============================================================================
SFS AND BYTE FILE COUNTER INFORMATION

       4440  File Copy Requests
          0  File Pool Control Backup Requests
    6463156  Get Directory Entry Requests
      30126  Get Directory Requests
      10070  Lock Requests
    2304753  Open Directory Requests
      14145  Open File New Requests
   11441250  Open File Read Requests
      55127  Open File Replace Requests
     361687  Open File Write Requests
      10174  Query Administrator Requests
```

Simple Exec

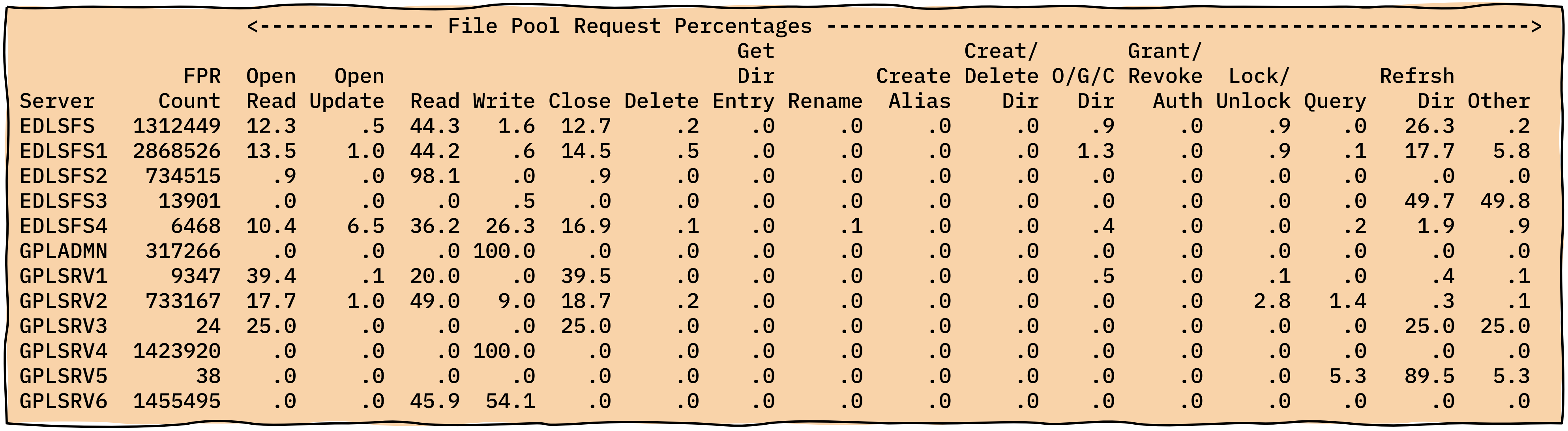
- Remember that counters are ‘running’ counters
- Can pick out some of our favorite counters and do simple math
- File pool requests per second is a good measure of the ‘work’ that SFS is doing.



```
Ready;
fpr-rate
0.299928437 File pool requests per second
Ready;
```

```
/* FPR-RATE Exec determines the filepool request rate */

'PIPE CMS Q FILEPOOL COUNTER',
'| locate /Total File Pool Requests/',
'| Spec w1 1',
'| VAR StartingFPR'

JUNK = TIME('R')

'CP SLEEP 10 SEC'

'PIPE cms q filepool counter',
'| locate /Total File Pool Requests/',
'| Spec w1 1',
'| VAR EndingFPR'

elapsed = TIME('R')

rrate   = (EndingFPR - StartingFPR)  / elapsed
say rrate  'File pool requests per second'

Exit
```

Performance Toolkit FCX152 SFSREQ Report



```
                  <-------------- File Pool Request Percentages -------------------------------------------------------->
                                                         Get               Creat/       Grant/
             FPR  Open   Open                            Dir        Create Delete O/G/C Revoke  Lock/       Refrsh
Server     Count  Read Update  Read Write Close Delete Entry Rename  Alias    Dir   Dir   Auth Unlock Query    Dir Other
EDLSFS   1312449  12.3     .5  44.3   1.6  12.7     .2    .0     .0     .0     .0    .9     .0     .9    .0   26.3    .2
EDLSFS1  2868526  13.5    1.0  44.2    .6  14.5     .5    .0     .0     .0     .0   1.3     .0     .9    .1   17.7   5.8
EDLSFS2   734515    .9     .0  98.1    .0    .9     .0    .0     .0     .0     .0    .0     .0     .0    .0     .0    .0
EDLSFS3    13901    .0     .0    .0    .5    .0     .0    .0     .0     .0     .0    .0     .0     .0    .0   49.7  49.8
EDLSFS4     6468  10.4    6.5  36.2  26.3  16.9     .1    .0     .1     .0     .0    .4     .0     .0    .2    1.9    .9
GPLADMN   317266    .0     .0    .0 100.0    .0     .0    .0     .0     .0     .0    .0     .0     .0    .0     .0    .0
GPLSRV1     9347  39.4     .1  20.0    .0  39.5     .0    .0     .0     .0     .0    .5     .0     .1    .0     .4    .1
GPLSRV2   733167  17.7    1.0  49.0   9.0  18.7     .2    .0     .0     .0     .0    .0     .0    2.8   1.4     .3    .1
GPLSRV3       24  25.0     .0    .0    .0  25.0     .0    .0     .0     .0     .0    .0     .0     .0    .0   25.0  25.0
GPLSRV4  1423920    .0     .0    .0 100.0    .0     .0    .0     .0     .0     .0    .0     .0     .0    .0     .0    .0
GPLSRV5       38    .0     .0    .0    .0    .0     .0    .0     .0     .0     .0    .0     .0     .0   5.3   89.5   5.3
GPLSRV6  1455495    .0     .0  45.9  54.1    .0     .0    .0     .0     .0     .0    .0     .0     .0    .0     .0    .0
```

- Log report, so actually shows these values over time
- Note “Percentages”
- Highlights the most important request types and then places the rest in the “Other” category

Agents

- Dispatchable tasks in the file pool server
- Typically associated with user work, mapping logical units of work (LUWs) in progress
- Number of agents determined by USERS value start-up parameter

$$\texttt{Number of Agents} = 4 + \texttt{Truncate}\left(\frac{\texttt{Users}}{8}\right)$$

- A single user can be associated with multiple agents if multiple workunits are active.
- If insufficient number of agents, work gets queued up. This is bad.
- If too many agents exist, memory may be wasted.
- Better to make number too big than too small
- Monitor ***Active Agents Highest Value*** in QUERY FILEPOOL or corresponding monitor data.

Monitoring Agents

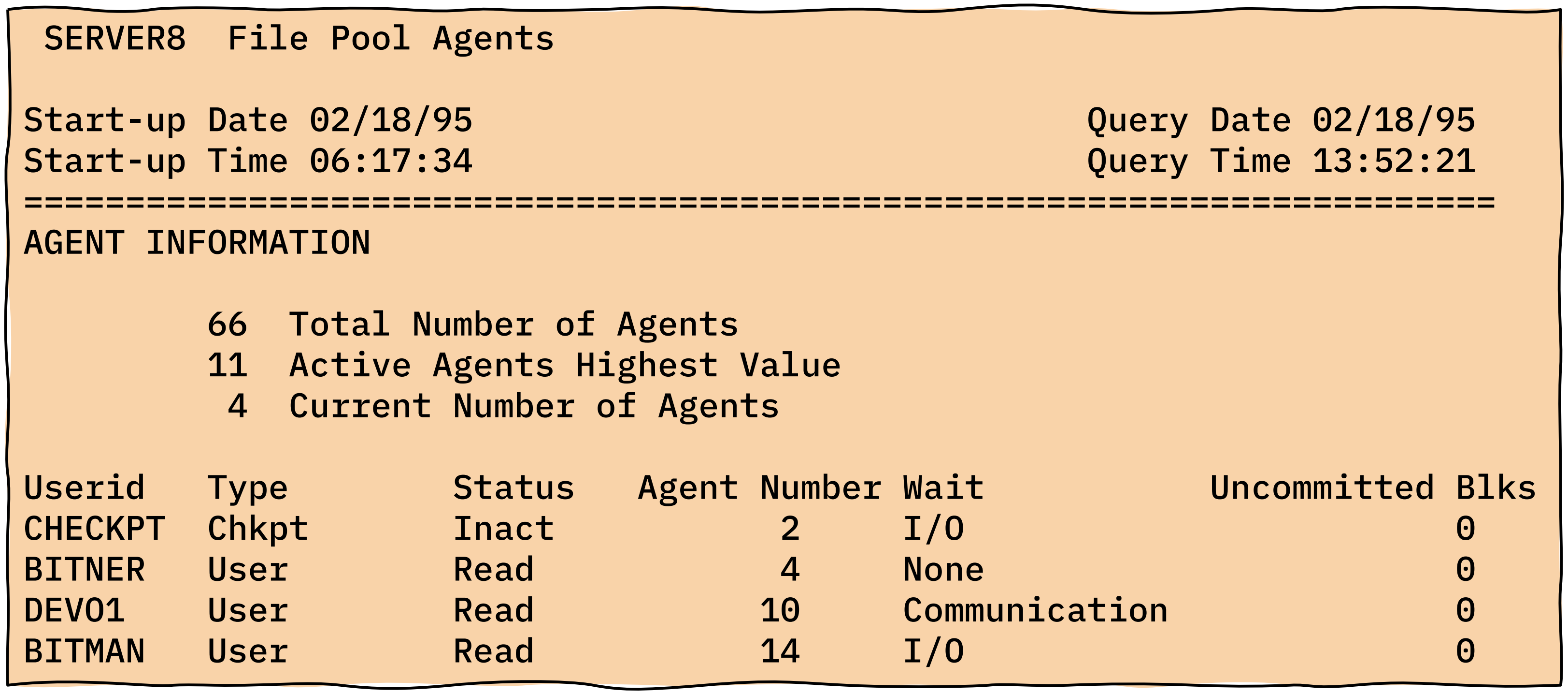
- Held Agents
  - Agents associated with a logical unit of work (LUW)
  - Agent Holding Time / Elapsed Time gives percentage of time Agents on mission
- Active Agents
  - Agents currently doing work on behalf of a file pool request
  - File Pool Request Service Time / Elapsed Time
- Typical ratio of held to active is < 10:1, though workloads vary
- For most workloads, applications remaining in an LUW without doing work is a warning sign.

QUERY FILEPOOL REPORT

- Example
  - Total: 79
  - Highest Active: 39 ☺
  - USERS parm: 600
  - 1 GB Virtual Machine

```
 EPLLIB1  File Pool Report

Start-up Date 03/14/21                              Query Date 04/07/21
Start-up Time 07:21:43                              Query Time 15:48:05
========================================================================
FILE POOL OVERVIEW INFORMATION

 1073741824  Virtual Storage Size in Bytes (   1048576 in KB)
     293308  Virtual Storage Highest Value in KB
          0  Virtual Storage Requests Denied
          0  Virtual Storage Reclaim Value

       2000  Maximum Number of Connections
        304  Connections Highest Value

         79  Total Number of Agents
         39  Active Agents Highest Value

        500  Maximum Number of Storage Groups
         12  Storage Groups in Use

        498  Maximum Number of Minidisks
        134  Minidisks in Use
```

QUERY FILEPOOL AGENT



```
 SERVER8  File Pool Agents

Start-up Date 02/18/95                              Query Date 02/18/95
Start-up Time 06:17:34                              Query Time 13:52:21
========================================================================
AGENT INFORMATION

         66  Total Number of Agents
         11  Active Agents Highest Value
          4  Current Number of Agents

Userid   Type        Status   Agent Number Wait           Uncommitted Blks
CHECKPT  Chkpt       Inact           2     I/O                        0
BITNER   User        Read            4     None                       0
DEVO1    User        Read           10     Communication              0
BITMAN   User        Read           14     I/O                        0
```

- Agent *Type* will be either “User” or some system type
- Various types of *Wait*

Performance Toolkit FCX150 SFSLOG

```
                        <-- Time per File Pool Request --->  <--- Server Utilization ---->  <--Agents--> Dead-
             FPR   FPR                    Block                           Page Check                     locks
Server     Count  Rate  Total   CPU  Lock   I/O   ESM Other  Total   CPU  Read point  QSAM  Active  Held w/ RB
EDLSFS   1312449 370.7   .001  .000  .000  .000  .000  .001    1.0    .9    .0    .0    .0      .2    .3     0
EDLSFS1  2868526 810.3   .001  .000  .000  .001  .000  .000    4.1   3.2    .0    .2    .7      .8   1.5     0
EDLSFS2   734515 207.5   .001  .000  .000  .001  .000  .000     .4    .4    .1    .0    .0      .2   1.0     0
EDLSFS4     6468   1.8   .001  .000  .000  .001  .000  .000     .0    .0    .0    .0    .0      .0   2.5     0
```

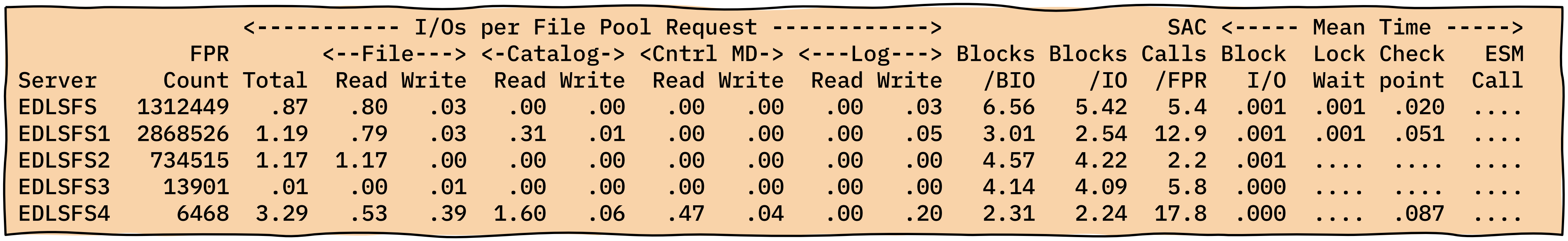
- Performance Toolkit does a lot of math for you!

- Time per File Pool Request is a breakdown of the time to process a file pool request

- Server Utilization speaks to the SFS file pool server virtual machine
  - QSAM refers to control data backup time

- This report also gives you the Active and Held Agent values.

Performance Toolkit FCX151 SFSIOLOG



```
                 <----------- I/Os per File Pool Request ------------>                 SAC <----- Mean Time ----->
             FPR       <--File---> <-Catalog-> <Cntrl MD-> <---Log---> Blocks Blocks Calls Block  Lock Check   ESM
Server     Count Total  Read Write  Read Write  Read Write  Read Write   /BIO    /IO  /FPR   I/O  Wait point  Call
EDLSFS   1312449   .87   .80   .03   .00   .00   .00   .00   .00   .03   6.56   5.42   5.4  .001  .001  .020  ....
EDLSFS1  2868526  1.19   .79   .03   .31   .01   .00   .00   .00   .05   3.01   2.54  12.9  .001  .001  .051  ....
EDLSFS2   734515  1.17  1.17   .00   .00   .00   .00   .00   .00   .00   4.57   4.22   2.2  .001  ....  ....  ....
EDLSFS3    13901   .01   .00   .01   .00   .00   .00   .00   .00   .00   4.14   4.09   5.8  .000  ....  ....  ....
EDLSFS4     6468  3.29   .53   .39  1.60   .06   .47   .04   .00   .20   2.31   2.24  17.8  .000  ....  .087  ....
```

- Breaks down I/O rates on per file pool request
- Shows the blocks per I/O
- These file pools not controlled by an External Security Manager (ESM), so no values there.

Checkpoint Processing

- Should normally be less than 4 seconds
  - Checkpoint Time / Checkpoints Taken

- Long checkpoint time affects mostly response time, not resource consumption

- Longer checkpoint time from
  - Too few control buffers
    - Control Minidisk Blocks Read / Total File Pool Requests should be less than 0.005
  - Poor I/O performance
  - Large number of changed catalog buffers

Catalogs

- Fragmented catalogs
  - Watch for increase in catalog blocks read per file pool request
  - Reorganize the catalogs using **FILESERV REORG** command

- Catalog buffers (**CATBUFFERS**)
  - Trade off between I/O and memory
  - Default is based on **USERS** start-up parameter
  - Catalog Blocks Read / Total DASD Block transfers should be between 0.20 and 0.25

Additional Pearls

- Control data backups should be structured to avoid prime shift if possible
  - Watch for spikes of high QSAM (serial I/O) time
  - Doing backups to another filepool is an alternative
  - Check size of log disks, increase size to lower the frequency of control data backups

- Use DMSFILEC or COPYFILE command to move data from one file to another file in the same filepool instead of copying to minidisk and then back to SFS.

SFS History

- VM/SP 6
  - It all started

- VM/ESA 1.1.0
  - Introduced some asynchronous function calls
  - Reduction in filepool requests per command
  - Improvements in locking to minimize rollbacks due to deadlock

- VM/ESA 1.1.1
  - VM data space support
  - Checkpoint improvements
  - Asynchronous function calls through CSL

- VM/ESA 1.2.0
  - More checkpoint improvements
  - Improved catalog insert algorithm

- VM/ESA 1.2.1
  - Improved control data backup
  - SFS thread blocking I/O

- VM/ESA 1.2.2
  - Improved handling of released file blocks
  - Performance improvements for Revoke

The End of Part II

References in z/VM Library

- CMS File Pool Planning, Administration, and Operation
- Performance
- CMS Application Development Guide
- CP Planning and Administration
- CP Command and Utility Reference
- Performance Toolkit Reference
- CMS Planning and Administration
- CMS Callable Services Reference