



OS/390 UNIX Performance and Tuning

Share Session 2567

Don Corbett
dpcorbet@us.ibm.com
08/23/99





OS/390 UNIX Performance and Tuning

Special Notices

This presentation was produced in the United States. IBM May not offer the products, programs, services or features discussed herein in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the products , programs, services and features available in your area. Any reference to an IBM product, program, service or feature is not intended to state or imply that only IBM's product, program, service or feature may be used. Any functionally equivalent product, program, service of feature that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program, service or feature.

IBM may have patents or pending patent applications covering subject matter in this presentation. The furnishing of this presentation does not give you any license to these patents. Send license inquiries, in writing, to IBM Director of licensing, IBM Corporation, 500 Columbus avenue, Thornwood, NY 10594 USA.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals or objectives only. Contact your local Branch Office or IBM Authorized Reseller for the full text of a specific Statement of General Direction.



OS/390 UNIX Performance and Tuning

Special Notices (cont.)

The information contained in this presentation has not been submitted to any formal IBM test and is distributed on an "as is basis" without any warranty either expressed or implied. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

The information contained in this document represents the current views of IBM on the issues discussed as of the date of publication. IBM cannot guarantee the accuracy of any of the information presented after the date of publication.

Any performance data contained in this presentation was determined in a controlled environment; and therefore, the results which may be obtained in other operating environments may vary significantly. Some measurements quoted in this presentation may have been made on development-level systems. There is no guarantee that these measurements will be the same on generally available systems . (continued on next page)

(C) Copyright IBM Corporation 1999

Don Corbett



OS/390 UNIX Performance and Tuning

Special Notices (cont)

Some of the measurements quoted in this presentation may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company, Limited. Other company, product and service names may be trademarks or service marks of others.

Trademarks or Registered trademarks of International Business Machines Inc.

ACF/VTAM tm
CICS tm
DB2 tm
DFSMS tm
ESCON tm
IBM tm
IMS tm

MVS tm
OpenEdition tm
OS/390 tm
RACF tm
RMF tm
S/390 tm



An OS/390 Performance Analyst's view

▶ What is OS/390 UNIX Services

- Also known as OpenEdition MVS (OE)

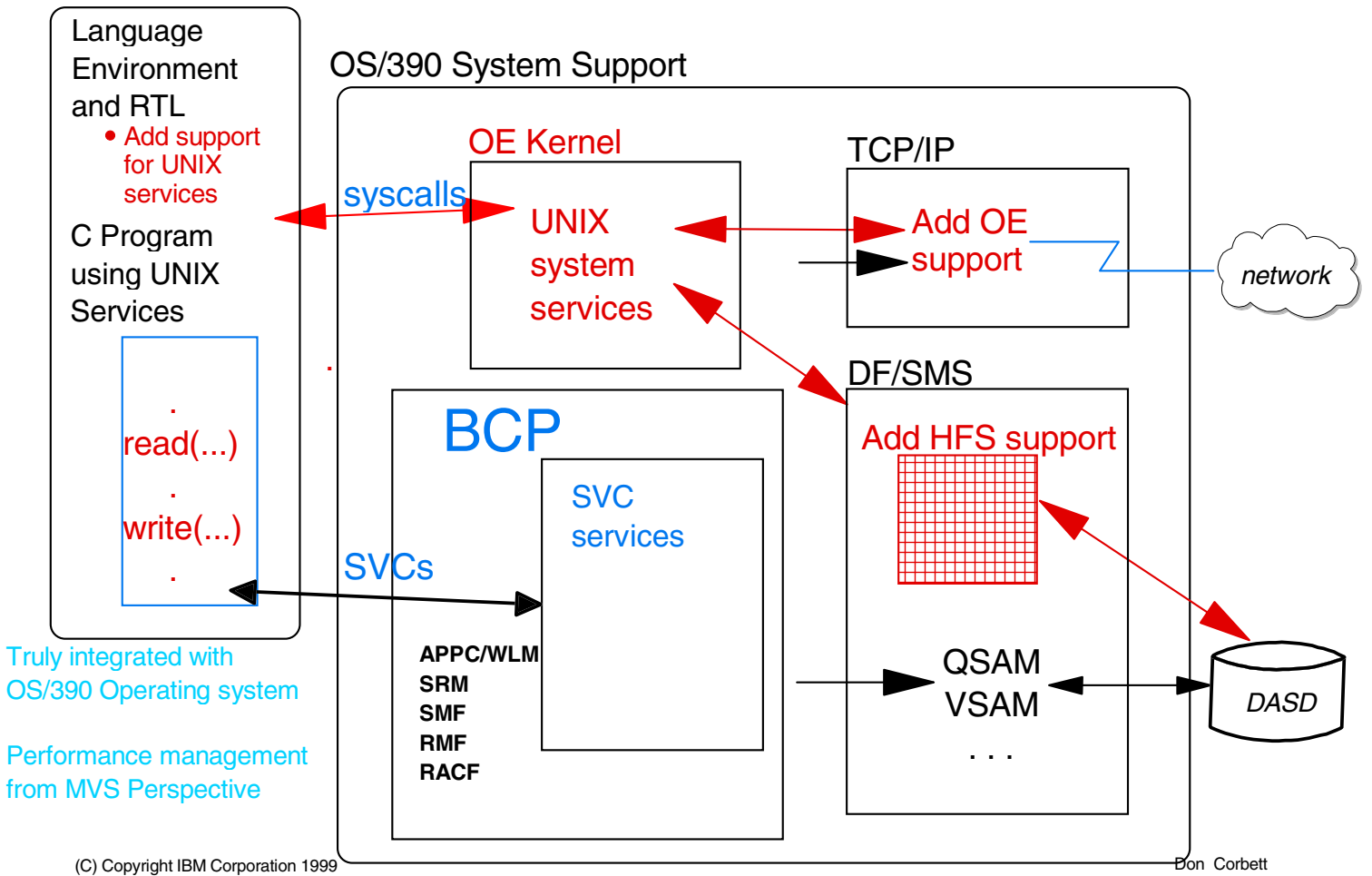
■ Performance Tuning Recommendations

■ Recent Performance Improvements



OS/390 UNIX Performance and Tuning

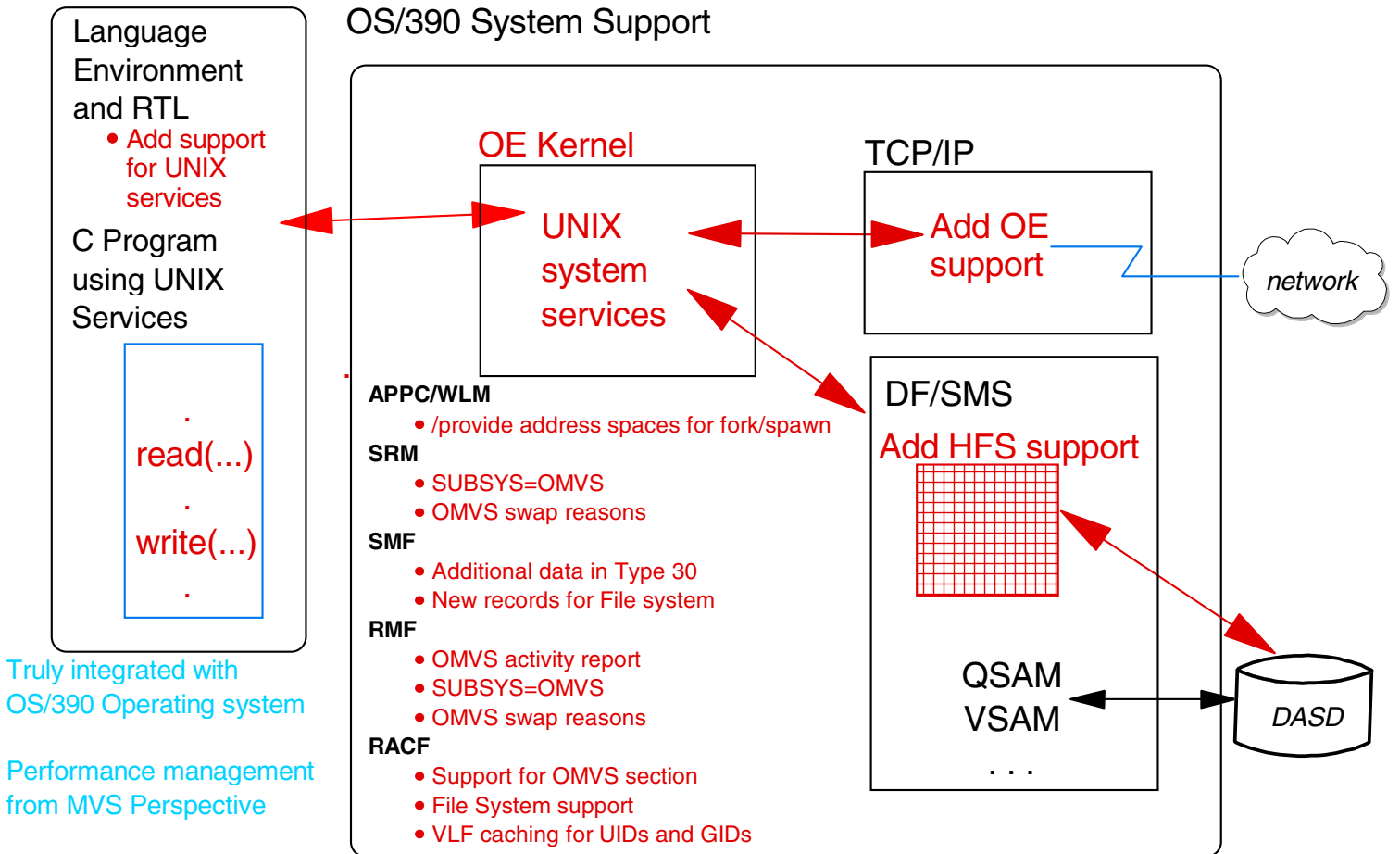
Portability of applications and programming skills





OS/390 UNIX Performance and Tuning

Portability of applications and programming skills



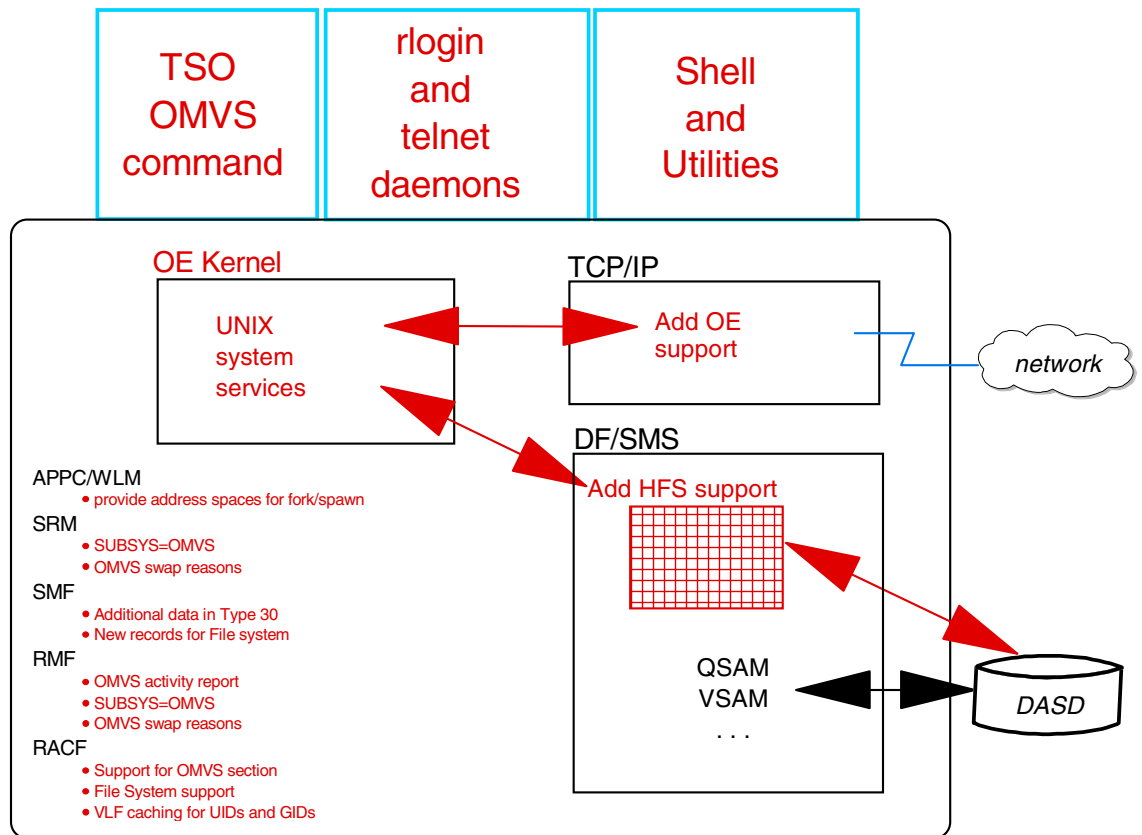
Truly integrated with OS/390 Operating system

Performance management from MVS Perspective



OS/390 UNIX Performance and Tuning

Application Porting Interactive Environment





OS/390 UNIX Performance and Tuning

Performance Monitoring, MVS style

CPU Activity Report

New Address space type:

- BATCH
- STC
- TSO
- APPC
- **OMVS**

Workload Activity Report

Performance Groups assigned for
SUBSYS=OMVS address spaces

Swap Placement Activity Report

New Swap Reasons:

- Terminal Input Wait
- Detected Wait
- Long Wait
-
- **OMVS Input Wait**
- **OMVS Output Wait**

OMVS Kernel Activity Report

- OE Syscalls / Second
- Monitor OE System Limits
 - Process Activity
 - ▶ processes on system
 - ▶ users on system
 - ▶ processes/user
 - Inter-Process Communication
 - ▶ message queue IDs
 - ▶ semaphore IDs
 - ▶ shared memory IDs
 - ▶ shared memory pages
 - Memory mapped files
 - ▶ memory map storage pages
 - Total Shared Storage pages



An OS/390 Performance Analyst's view

- What is OS/390 UNIX Services
 - Also known as OpenEdition MVS (OE)
- ▶ Performance Tuning Recommendations
- Recent Performance Improvements



OS/390 UNIX Performance and Tuning

Unless otherwise noted all performance results contained in this presentation are from IBM internal workload measurements.



Tuning for OE is a MUST!

Large "make"

Untuned - 70 minutes

Tuned - 14 minutes

(5X better!)

Slow "ls -l"

Untuned - 30 seconds

Tuned - 2 seconds

(15X better!)



Key Tuning Considerations

- Central Storage
- Runtime Routines in LPA
- RACF Caching of UIDs and GIDs
- APPC Initiators
- File System Organization
- Use of heappools LE option
- Use of Arch() compile option



Ensure Enough Central Storage

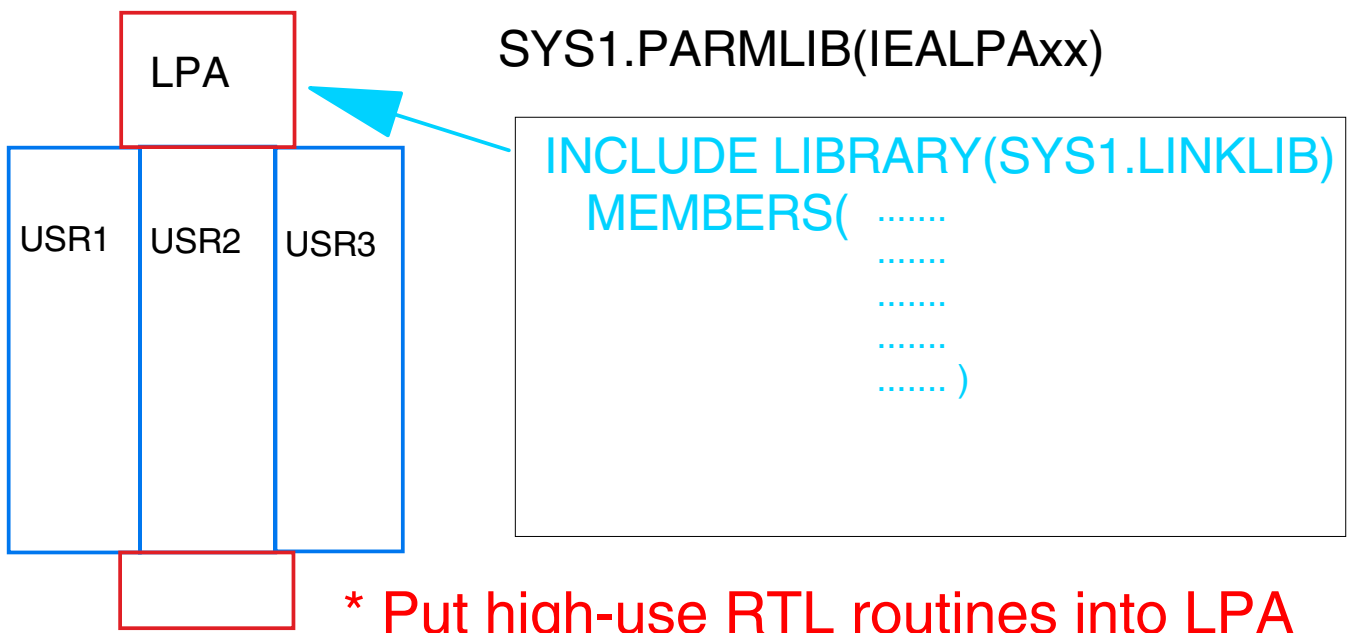
Small LPAR or VM Guest

Minimum - 64 Meg

More needed if...

- Fast / frequent access to large files
- Many concurrently active users
- Storage intensive applications

Runtime Routines in LPA



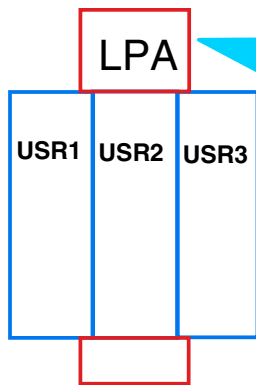
*** Put high-use RTL routines into LPA**

- common storage for all address spaces
- reduced DASD I/O for loads
- less storage copied on fork
- better response time and throughput



OS/390 UNIX Performance and Tuning

Runtime Routines in LPA, OS390 R4 - R7



Below the line LPA storage
IEFIB600 (45k)
IEWBLINK (2k)
CEEINIT (48k)

```
SYS1.Parmlib(LPALSTxx)
```

```
CEE.SCEELPA
```

```
SYS1.PARMLIB(IEALPAxx)
```

```
INCLUDE LIBRARY(SYS1.LINKLIB)
```

```
MEMBERS( IEFIB600, IEFXB603,  
         IEWBLINK, HEWL, HEWLDRGO,  
         HEWLH096, HEWLOAD, HEWLOADR,  
         IEWBLDGO, IEWBLOAD, IEWBLODI,  
         IEWBODEF, IEWL, IEWLDRGO,  
         IEWLOAD, IEWLOADI, IEWLOADR,  
         LINKEDIT, LOADER)
```

```
INCLUDE LIBRARY(CBC.SCBCCMP)
```

```
MEMBERS( CBCDRVR, CBC30PTP, CBC30,  
         CBC3P, CBC3T)
```




OS/390 UNIX Performance and Tuning

OS/390 V2R4 or later

- ✓ Need more than one LE level?
- ✓ Can't put current LE modules in LPA?
- ★ Use RTLS rather than STEPLIB

SYS1.PARMLIB(BPXPRMxx)

```
...  
RUNOPTS('RTLS(ON) LIBRARY(SYSCEE)')  
...
```

SYS1.PARMLIB(CSVRTLxx)

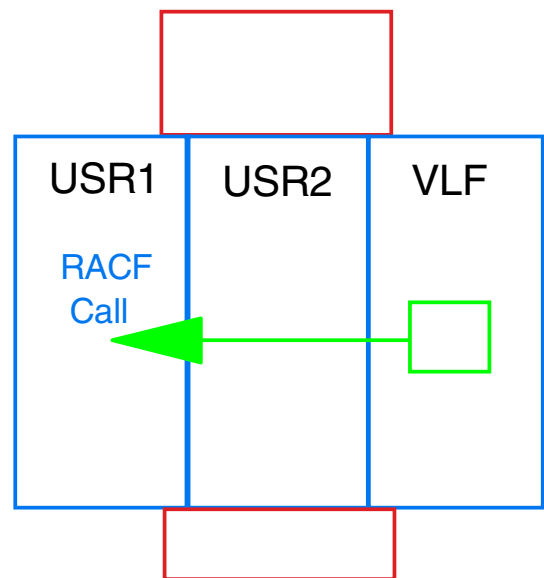
```
PHYSICAL (  
    LIBRARY(SYSCEE) ADD  
    DSLIST(CEE.SCEERUN)  
    MAXABOVEP(10M)  
    MAXBELOWP(50K)  
)  
LOGICAL (  
    LIBRARY(SYSCEE) VERSION(1) DEFAULT  
    PHYSICAL(SYSCEE)  
)
```

RACF Caching of UIDs and GIDs

SYS1.PARMLIB(COFVLFxx)

```

.....
.....
CLASS NAME(IRRUMAP)
  EMAJ(UMAP)
CLASS NAME(IRRGMAP)
  EMAJ(GMAP)
CLASS NAME(IRRGTS)
  EMAJ(GTS)
CLASS NAME(IRRACEE)
  EMAJ(ACEE)
  
```



★ Ensure that RACF caches UID and GID information in VLF (don't forget to start VLF)

- reduced DASD I/O for RACF calls
- reduced CPU searching through all userids and groups
- faster response time

★ Make sure all HFS files have valid UIDs and GIDs!!



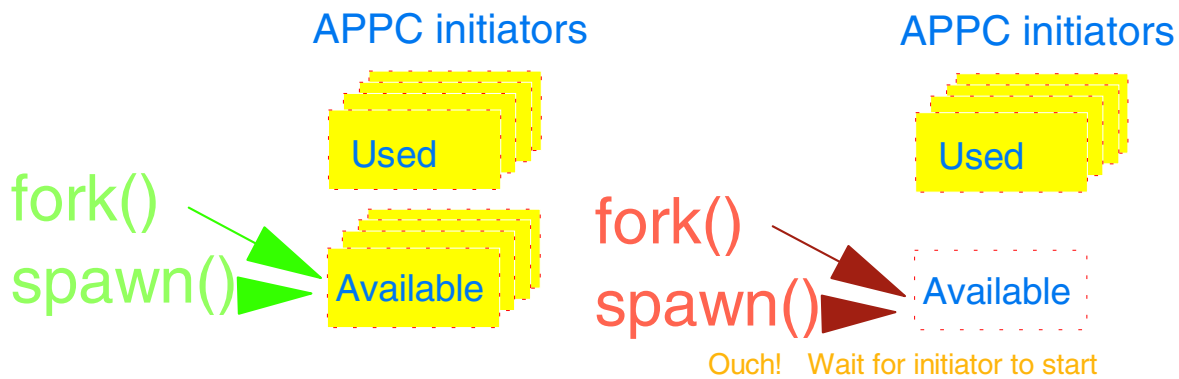
OS/390 UNIX Performance and Tuning

APPC and WLM Initiators

SYS1.PARMLIB(ASCHPMxx)

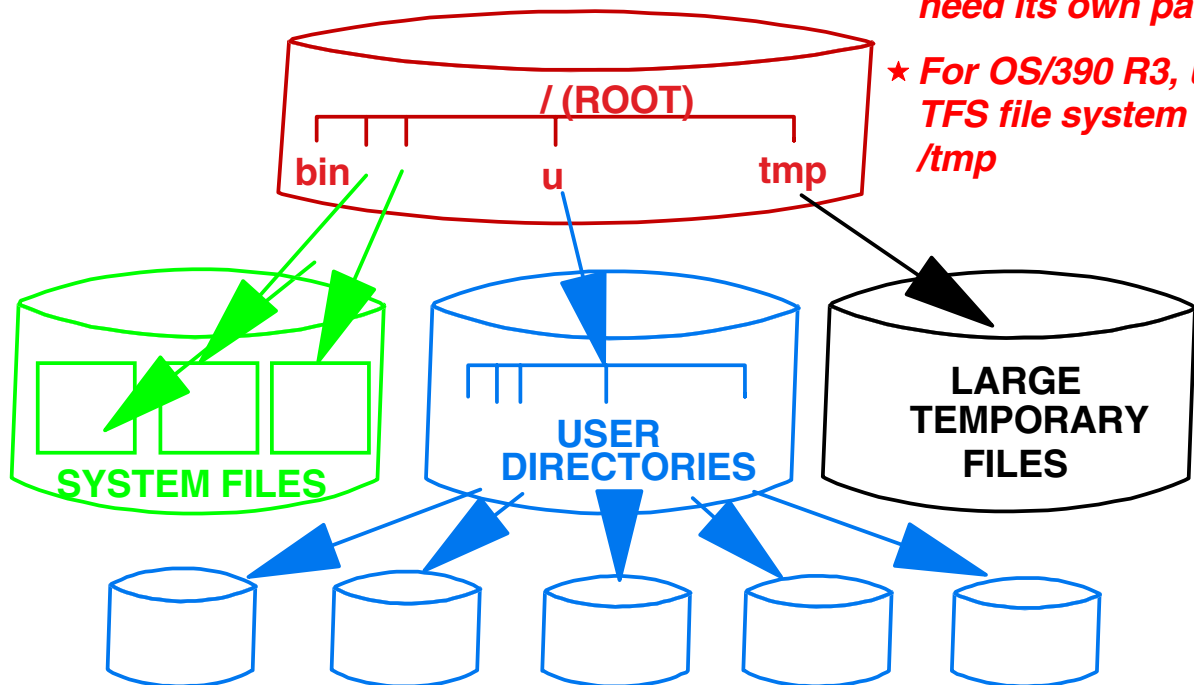
```
CLASSADD CLASSNAME(OPENMVS)
MIN(20) MAX(300) RESPGOAL(1)
```

Note: This is no longer needed on OS/390 V2R4



- ★ Set MIN high enough to avoid constant start and stop of initiators
 - with fewer users, initiator requirements can fluctuate wildly
- ★ But, setting MIN too high can waste storage resources
- ★ Set MAX to what your system can tolerate
 - not more than the MAXPROCSYS value in BPXPRMxx

File System Organization



★ *If lots of interactive users, /tmp may need its own pack*

★ *For OS/390 R3, use TFS file system for /tmp*

★ **Define separate HFS files for users or groups**

- will reduce contention
- D GRS,C from MVS console shows contention

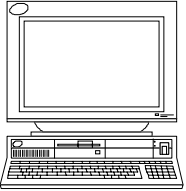
★ **CU caching with DASD Fast Write helps**

- have seen significantly faster makes due to DFW



OS/390 UNIX Performance and Tuning

SUPERUSER



OS/390 V2R4 or later

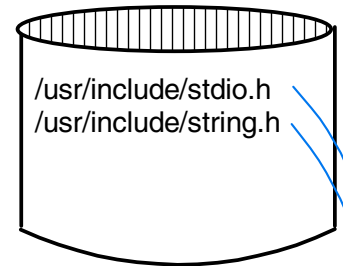
★ Use the filecache command to cache read-only files in the OE Kernel

/etc/rc

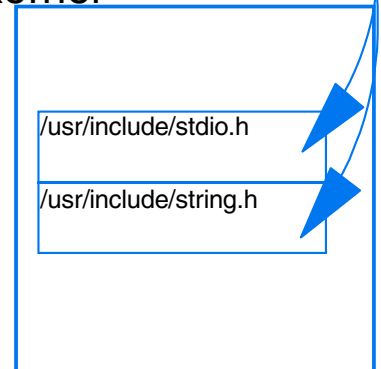
```
...  
/usr/sbin/filecache -a /usr/include/stdio.h  
/usr/sbin/filecache -a /usr/include/string.h  
...
```

- Read-only files
 - If file is updated, must remount filesystem to put file in filecache
- Major reduction in CPU to read file
 - 33% to 50% reduction (informal tests)
- Major reduction in Elapsed time to read file
 - No I/O once the file is in the Kernel

HFS



Kernel





HFS in DFSMS 1.5

- Deferred Writes
- Sync Parm on File mount default 60 seconds
- Virtual - size of HFS buffer pool in MB default 50% of real storage
- Fixed- page-fixed buffer pool size default 0 If HFS is not used no storage is allocated. For systems dedicated to HFS usage recommend Fisex=virtual up to 50% of real storage capacity.



DFSMS 1.5 RMF and Confighfs Reports

- Buffer pool hit ratios
- Frequency of buffer accesses
- Effectiveness of fixed buffers



Parms for FILESYSTYPE in BPXPRMxx

FILESYSTYPE TYPE(HFS)

ENTRYPOINT(GFUAINIT)

PARMS('SYNCDEFAULT(30)

VIRTUAL(128)

FIXED(128)')



Further improvements for the OpenEdition Development Environment

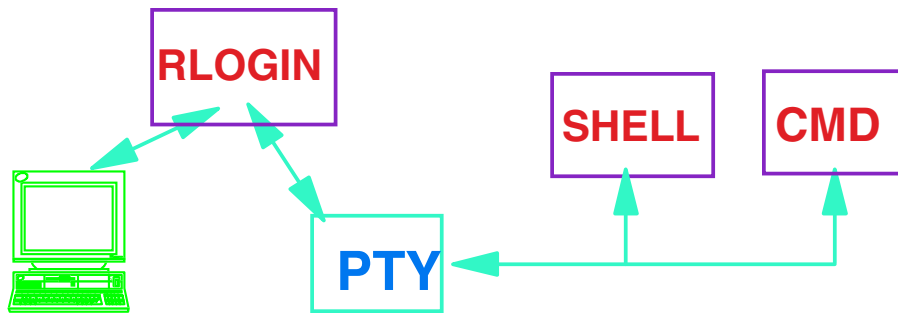
- Setting Key Environment Variables
- Avoiding STEPLIB propagation
- Putting c89, cc, and cxx in LPA



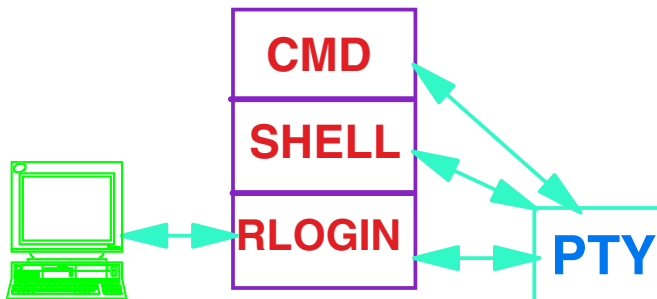
OS/390 UNIX Performance and Tuning

Use Environment Variables to Improve Performance

OE command with `_BPX_SHAREAS=NO`



OE command with `_BPX_SHAREAS=YES`

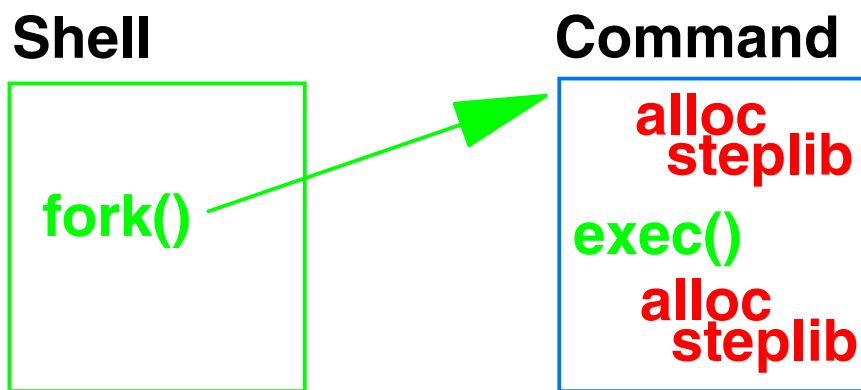


★ **Set `_BPX_SHAREAS=YES`**

– add to `/etc/profile`

– export `_BPX_SHAREAS=YES`

Prevent Propagation of STEPLIBs



- STEPLIB dynamically allocated for every fork, exec, spawn
- STEPLIB concatenation searched for all loads

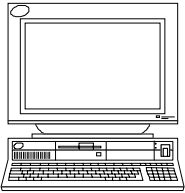
**** Add the code below to /etc/profile or \$HOME/.profile***

```
if [ -z "$STEPLIB" ] && tty -s;
then
    export STEPLIB=none
    exec sh -L
fi
```



OS/390 UNIX Performance and Tuning

SUPERUSER

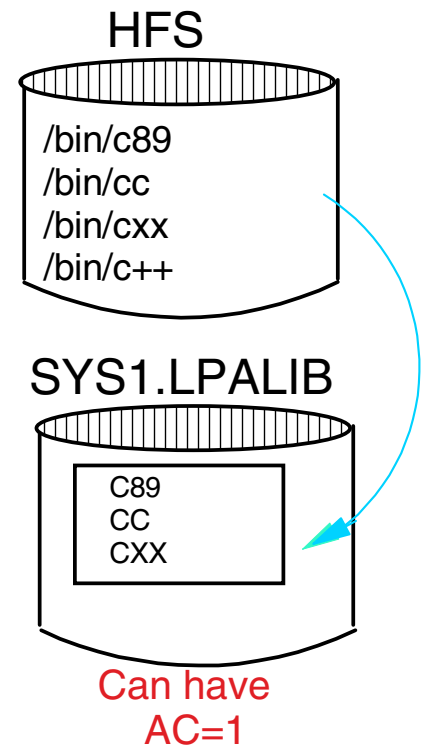


★ Put c89, cc, cxx in LPA

1. Re-linkedit the executables into an MVS partitioned dataset
2. Put Partitioned dataset into LPALST
3. Use chmod to turn on the sticky bit
4. Make c++ a symlink to cxx

- Saves main storage if multiple concurrent compiles
- Avoids load from disk
- prevents local spawn
 - c89, cc, cxx, c++ will fork if locally spawned

Reverts back to HFS if not found in LPA or LNKLST





OS/390 UNIX Performance and Tuning

★ Put c89, cc, cxx in LPA

JCL to Linkedit the c89

```

//PUTIN PUA JOB MSGLEVEL=(1,1)
//LINKIT EXEC PGM=IEWL,REGION=4M,
// PARM='LIST,XREF,LET,RENT,REUS,AMODE=31,RMODE=ANY,CASE=MIXED'
//SYSUT1 DD UNIT=SYSDA,SPACE=(CYL,(10,10))
//SYSPRINT DD SYSOUT=*
//INLMOD DD PATH='/bin/'
//SYSLMOD DD DSN=OECMD.LPALIB,DISP=SHR
//SYSLIN DD *
INCLUDE INLMOD(c89)
ALIAS CC
ALIAS CXX
ENTRY CEESTART
NAME C89(R)

```

1. Linkedit the commands into an MVS PDS

2. Add the commands to LPALST

Add PDS to

```

SYS1.PARMLIB(LPALSTxx)

```

```

...
OECMD.LPALIB

```

3. Turn on the sticky bits

Turn on the sticky bits

```

chmod +t /bin/c89 /bin/cc /bin/cxx

```

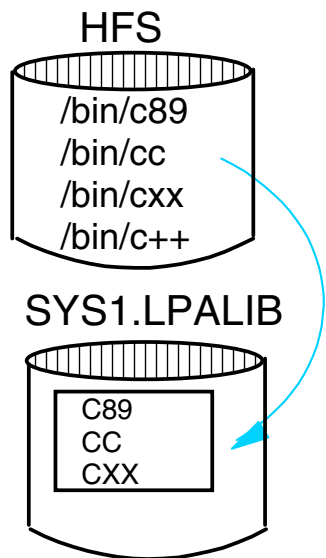
4. Make c++ symlink to cxx

Make c++ a symlink to cxx

```

mv /bin/c++ /bin/c+++
ln -s /bin/cxx /bin/c++

```



SUPERUSER



Don Corbett



Other Tuning Steps

- Tune SRM Parameters in Parmlib
- Tune OE limits in BPXPRMxx Parmlib member



OS/390 UNIX Performance and Tuning

Tune SRM Parameters in Parmlib

1. VTAM highest priority alone
2. TCP/IP 2nd priority alone
3. Put OMVS Kernel in high priority STC PGN
4. OS/390 R3 and later, put OE Init Process BPXOINIT in high priority STC PGN
5. Put SYSBMAS address space in high priority STC PGN
6. Define multiple performance periods for OMVS forked children
7. Put INIT process and other daemons at higher priority than other forked children



OS/390 UNIX Performance and Tuning

Tune SRM Parameters in Parmlib

SYS1.PARMLIB(IEAICSxx)

SUBSYS=STC,PGN=10

TRXNAME=OMVS,PGN=12 /* OpenEdition Kernel */

TRXNAME=BPXOINIT,PGN=12 /* OE Init Process, R3 */

TRXNAME=SYSBMAS,PGN=12 /* DF/SMS BMF */

SUBSYS=OMVS,PGN=30 /* Forked Children */

USERID=OMVSKERN,PGN=40 /* Init Daemons */

SYS1.PARMLIB(IEAIPSxx)

DMN=40,CNSTR=(1,5)

PGN=12,(DMN=12,DF=F80)

PGN=30,(DMN=30,DP=F53,DUR=5K) /* P1 Forked Children */

(DMN=31,DP=F51,DUR=10K) /* P2 Forked Children */

(DMN=32,DP=M4) /* P3 Forked Children */

PGN=40,(DMN=40,DP=F60)



Tune OE Limits in Parmlib

SYS1.PARMLIB(BPXPRMxx)

Control Process Activity

- ▶ MAXUIDS
- ▶ MAXPTYS
- ▶ MAXRTYS
- ▶ MAXPROCSYS
- ▶ MAXPROCUSER

Control the use of ESQA storage

- ▶ MAXSHAREPAGES
 - shared pages used by shared memory, memory map files, ptrace, fork (with COW)
- ▶ FORKCOPY
 - If run-time library is in LPA, specify "COPY"
 - If run-time library is not in LPA, specify "COW"



OS/390 UNIX Performance and Tuning

For More Information on Performance Tuning

Available On The Web

- Performance and tuning information
<http://www.s390.ibm.com/unix/bpxa1tun.html>
- Porting Guide
<http://www.ibm.com/s390/unix/bpxa1por.html>
- UNIX services publications
<http://www.ibm.com/s390/unix/bpxa1pub.html>
- Webmaster
<http://www.ics.raleigh.ibm.com/ics/icftopps.html>



OS/390 UNIX Performance and Tuning

For More Information on Performance Tuning

Books

- OS390 R1
 - SC28-1890-00 OS/390 OpenEdition MVS Planning
 - SG24-4529 MVS/ESA SP5.2.2 OpenEdition MVS Installation and Starter Kit
- OS390 R2
 - SC28-1890-01 OS/390 V1R2.0 OpenEdition Planning
- OS390 R3
 - SC28-1890-02 OS/390 V1R3.0 OpenEdition Planning
- OS390 R4
 - SC28-1890-0X OS/390 V2R4-5 OpenEdition Planning (where X is the release number-1)
- OS390 R6-R7
 - SC28-1890-XX OS/390 V2R6-7 UNIX Planning (where XX is the release number -1)
- See Web page
 - <http://ibm.com/s390/unix/bpxa1pub.html>

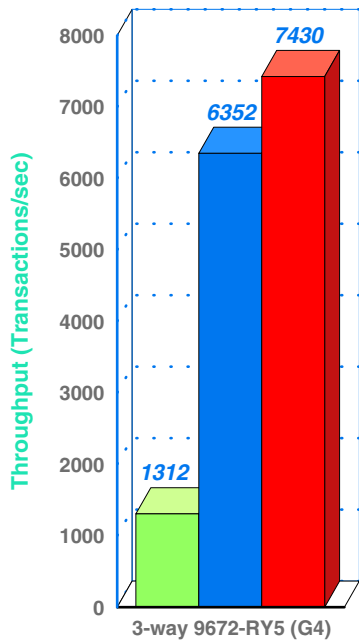


An OS/390 Performance Analyst's view

- What is OS/390 UNIX Services
 - Also known as OpenEdition MVS (OE)
- Performance Tuning Recommendations
- ▶ **Recent Performance Improvements**

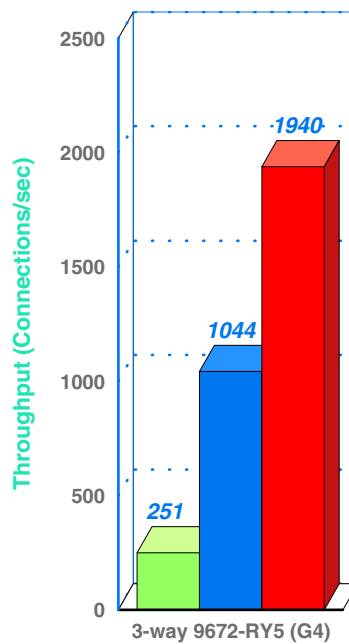
Performance Results

(Primitive Workloads - TCP/IP)



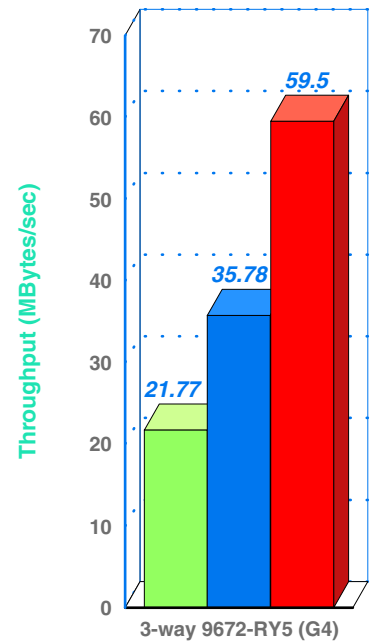
Request-Response Workload

5.66X Performance



Connect-Request-Response Workload

7.73X Performance



Stream Workload

2.73X Performance



TCP/IP V3R2 (9/96)



OS/390 V2R5 (3/98)

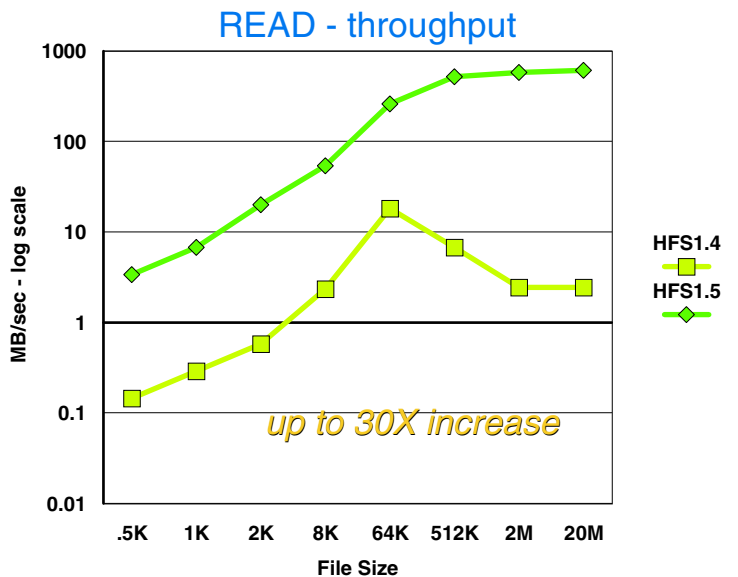
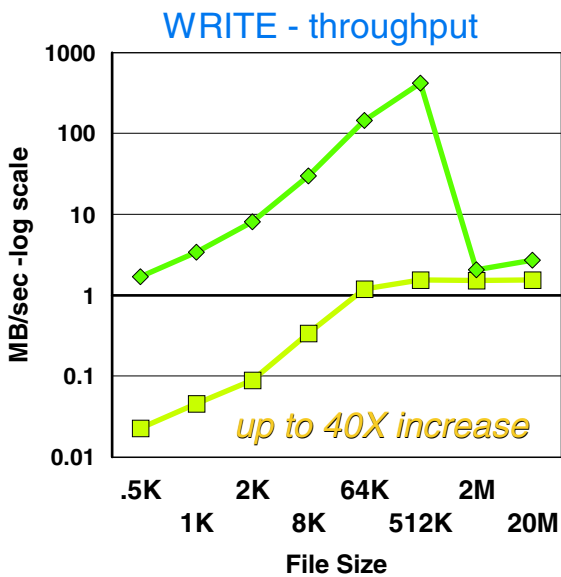


OS/390 V2R6 (9/98)

HFS Performance Restructure

DFSMS1.5 GA- 3/99

Runs on OS/390 R5,R6,R7

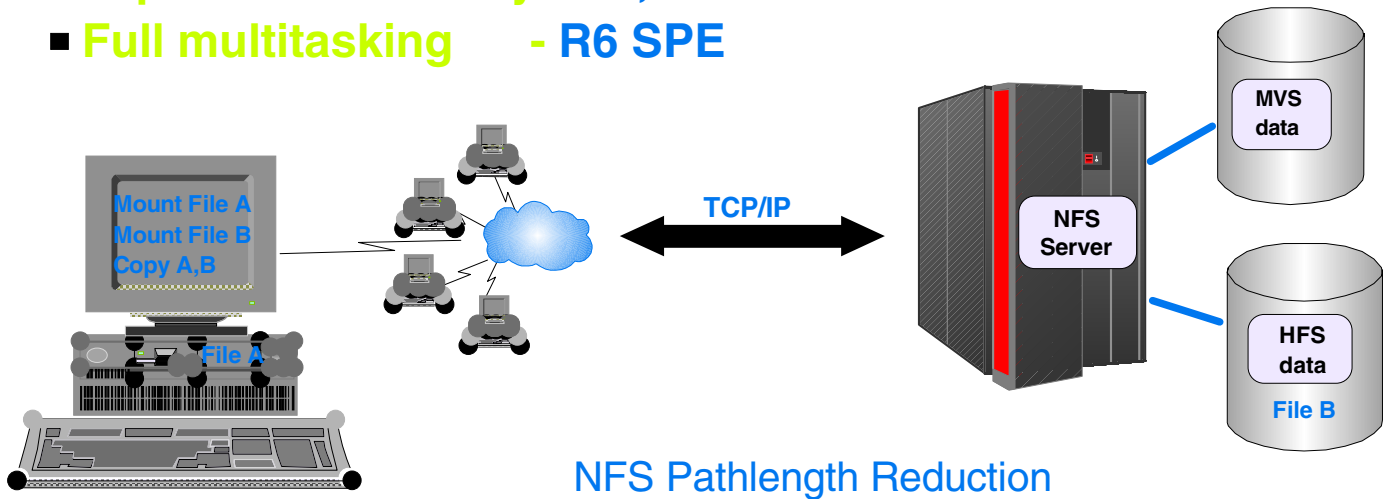


- Up to 85% reduction in CPU time
- Order of magnitude improvement in throughput for files < 512KB
- Optimized for multi-user access
- Make time improved by 75%

WRITE - cpu time

NFS Performance Improvements

- Pathlength reduction - R6,R7
- Improved Scalability - R6,R7
- Full multitasking - R6 SPE



NFS Pathlength Reduction

OS/390 Rel	R3	R6	R6(SPE)
Pathlength	Base	80%	90%

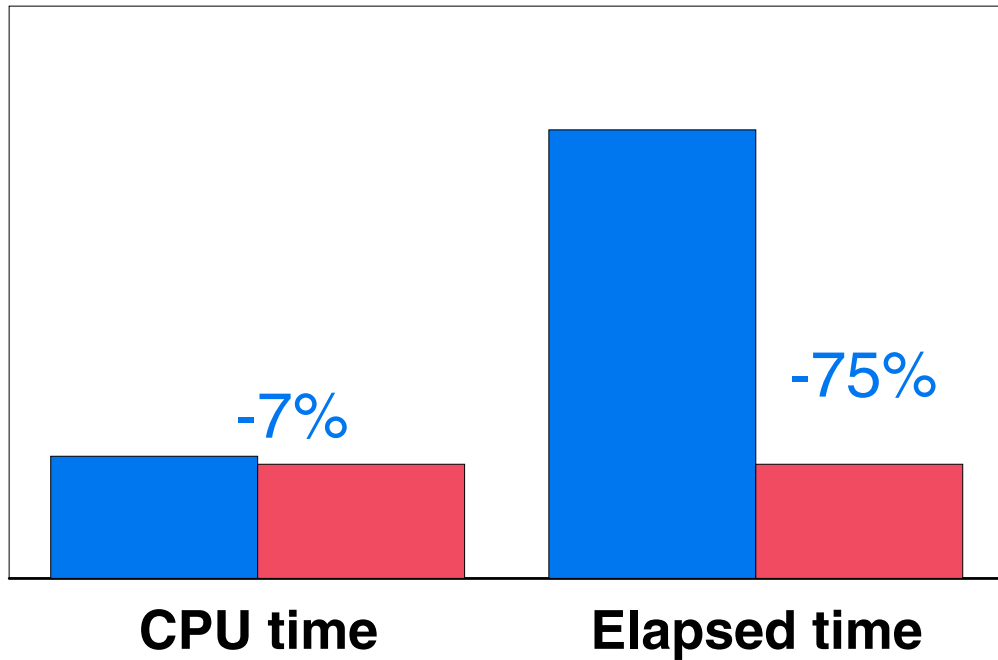
NFS Throughput Gains - R6 only

File Size	48K	256k	20MB	100MB
Single Client Read	55%	70%	85%	30%
Single Client Write	70%	10%	5%	30%
Multi-client Read	*	*	180%	*
Multi-client Write	*	*	50%	*

* not measured



Make test





Key Tuning Considerations Summary

- Central Storage
- Runtime Routines in LPA
- RACF Caching of UIDs and GIDs
- Initiators
- File System Organization
- Setting Key Environment Variables
- Avoiding Steplib propagation
- Putting c89,cc,and cxx in LPA
- SRM parameters in parmlib
- UNIX services limits in BPXPRMxx Parmlib member
- Use heappools to improve malloc performance
- Use of the Arch() compile option