IBM

Redbooks Paper

Nin Lei
Duane Mauler
J B Mills
John Mullin
Gene Ong
Judy Viccica

Linux on zSeries: Samba-3 Performance Observations

This Redpaper was developed to provide readers with basic performance guidelines for Samba Version 3.0.5 on zSeries® Linux® and to give background information that can be used in configuring and tuning the Samba V3 environment.

The data collected was developed using an IBM internal workload generator to simulate a heavy transaction load. The paper identifies parameters for optimum performance of Samba V3.

The system configuration

The test system was a zSeries 990 configuration with a FICON®-attached Enterprise Storage System (ESS 800). Table 1 summarizes the system configuration.

Table 1 Summary of the system configuration

Server model	CPCs and CPUs	Mode	Storage	System
zSeries 2084 (z990)	1 LPAR up to 4 CPs	Native	1 GB Main Memory per LPAR ESS 800 24 3390 Mod 9 Volumes	SuSE Linux SLES 8 SP3 2.4.21-241 Kernel Samba V3

The network interface

An Ethernet network was used to connect the zSeries Samba V3 server to the Client xSeries® x335 workstations that executed the test application. Table 2 summarizes the network interface.

Table 2 Summary of the network interface

Transport	Type	NIC	Code level	Switch
TCP/IP	Ethernet	OSA Express Gigabit Ethernet MTU=1500	QDIO base support level Rev 1.145.4.5 Rev 1.66.4.3 level 0615 QETH Rev 1.337.4.22 Rev 1.113.4.8	Cisco 6509

The hardware configuration

Figure 1 shows the hardware environment for the testing when one OSA card was enabled.

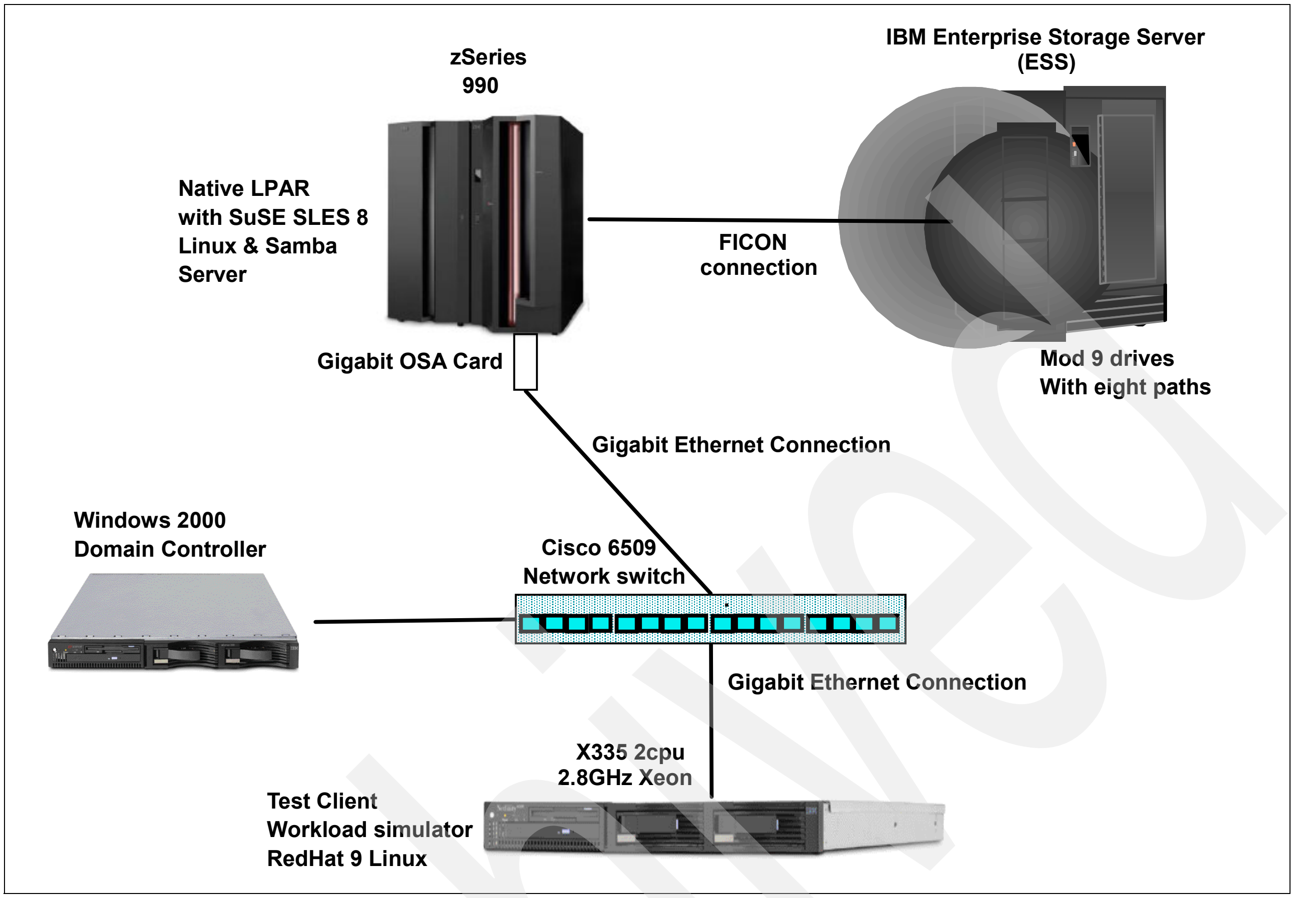

Figure 1 Single OSA interface card configuration

Figure 2 on page 4 shows the hardware environment when multiple OSA interface cards were enabled.

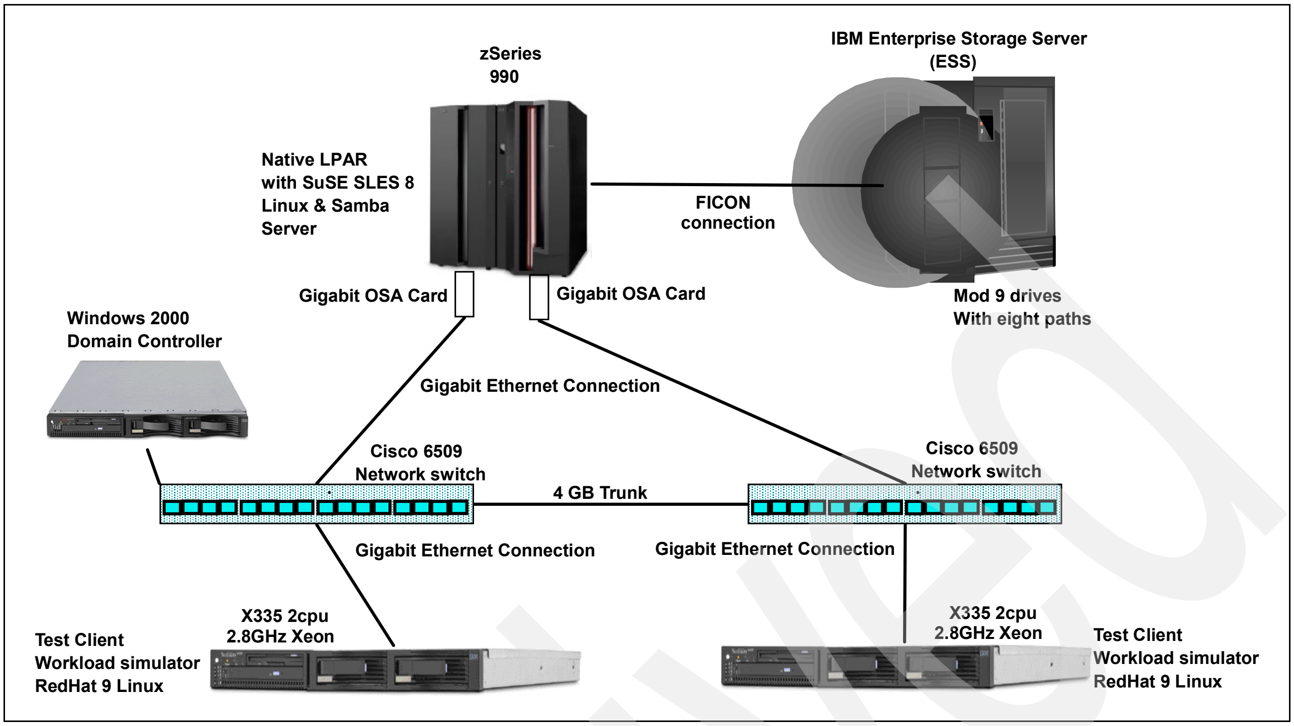

Figure 2 Multiple OSA interface cards configuration

Configuration, setup, preliminary tests, and the challenges

In the following sections we describe the configuration, workload driver, Samba configuration parameters, and tests that we performed.

System configuration

The test environment was a native LPAR on a zSeries 990 system. SUSE SLES 8 31-bit Linux distribution with Service Pack 3 was installed in the LPAR to provide the execution platform. Samba Version 2.2.8, the default version packaged with SUSE SLES 8, was used to run a number of tests to serve as a baseline measurement against Version 3.0.5. Then Samba Version 3.0.5 was manually downloaded, compiled and configured. The majority of measurements taken were on Samba Version 3.0.5.

The Linux kernel was upgraded to the 2.4.21-241 level. The kernel was changed to take advantage of upgrades that provide better cache performance. The 2.4.21-241 version of the kernel was used for the remaining test cases.

Workload driver

A mix of small, medium, and large files were hosted in the Samba share. An IBM internal workload simulator was used to generate user processes that accessed the files in the share and to produce a load to exercise the network and CPU(s). For each of the test runs, the run accessed a combination of 12,000 individual files and directories in the share. These files were distributed in the share to mimic a real-life data repository. For each run, the files were deleted and then recreated with the proper size for the test case. The load consisted of an environment with the read/write mix being 80% read and 20% write for all tests.

The workload simulator also allowed a “think time” between each access of a file in the share. We wanted to stress the system so a short think time was chosen (0.1 seconds), which provided rapid access of the files. The two major variations for the test runs were the number of users and the file size. Each test run had a duration of 600 seconds. The run was allowed to stabilize for three minutes before utilization sampling began.

Samba configuration parameters

During the configuration of Samba, two changes to the smb.conf parameters that may enhance performance are lowering the max xmit, and specifying the send and receive buffers of socket options. In this case, no difference was observed in changing the max xmit. Other tests have shown that max xmit = 8192 can improve performance.

The values that were used for the final runs were:

```
max xmit = 65536
socket options = TCP_NODELAY IPTOS_LOWDELAY SO_SNDBUF=14596 SO_RCVBUF=14596
```

The max xmit sets the maximum size of packets that Samba exchanges with a client. The socket options are host system tuning parameters set on a per-connection basis. Terms include:

TCP_NODELAY Have the server send packets as fast as possible to keep delay low.

IPTOS_LOWDELAY Used in conjunction with TCP_NODELAY to keep delay low.

SO_SNDBUF and SO_RCVBUF Sets the size of the send and receive buffers.

Several test runs were executed with various Samba log levels. A log level of 0 is recommended for any production environment. For reference, the smb.conf file used for the testing is included in “Appendix A” on page 11.

Preliminary test case 0.1 - ext2 or ext3 file system

The majority of the test cases were executed using the ext3 file system. All tests used the default journal parameters. Test runs were made using the ext2 file system to verify there is no significant CPU increase if ext3 with journaling is used.

Preliminary test case 0.2 - LVM and LVM stripe size

The Logical Volume Manager (LVM) was used to combine eight 3390-9 disks into a volume group. A striped logical volume was created from the volume group for the Samba share. Then the 12,000 files and directories were generated for each test. However, test runs were made with a small number of files to assess the effect of a file system without LVM.

From these runs, it appeared the overhead of LVM was not excessive and the advantages of LVM outweighed its impact on CPU usage; see Figure 3. Obviously, when the number and size of files surpass the capacity of a single available drive, LVM is necessary to provide capacity for the larger file systems.

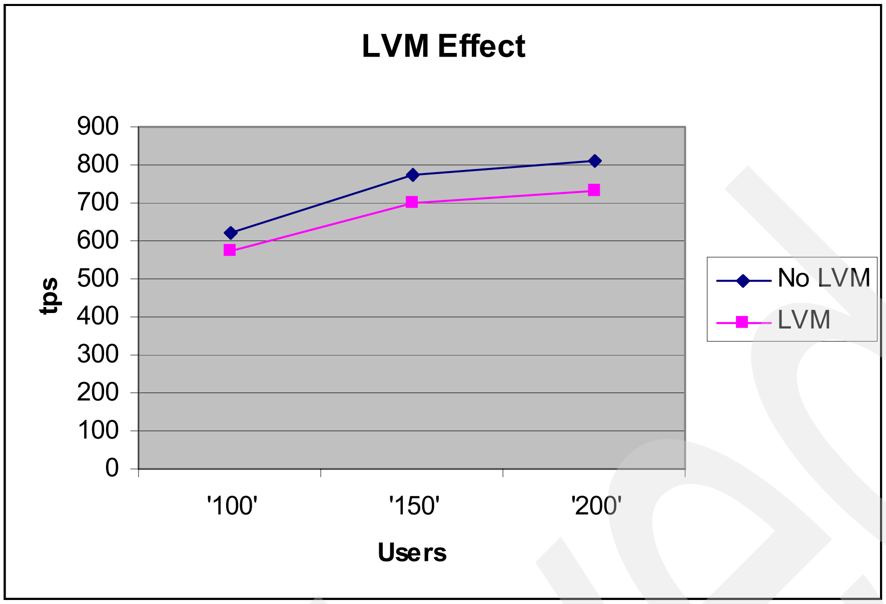

Figure 3 Effect of using Logical Volume Manager

A 4K LVM stripe size was used across the eight drives for the initial tests. Test runs were made with a varying number of users and file sizes. It was found that a stripe size of 64k yielded the optimal throughput; see Figure 4.

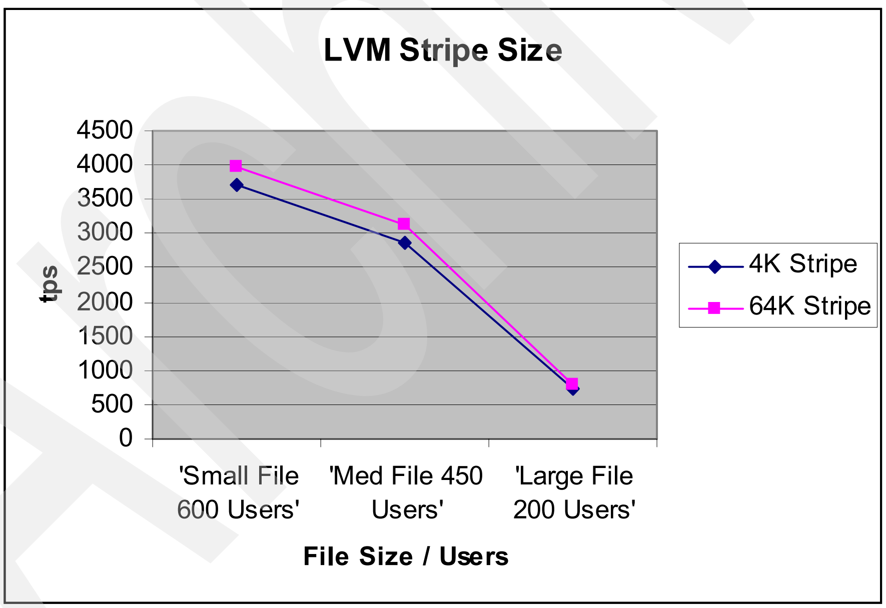

Figure 4 Effect of Logical Volume Manager stripe size

All tests used a Gigabit Open System Adapter (OSA) card to provide an Ethernet connection to the network. Using a large file size and ramping up the number of users, the tests were able to drive the OSA card to an extremely high percentage of use, while still having CPU resource available with one CPU.

A second OSA card was configured and the tests repeated using two CPUs and doubling the number of users. Again, the capacity of both OSA cards were exhausted before the CPU resources were consumed.

One area of note was uncovered in the configuration of the second OSA card. Samba will discover and use the network resources for outgoing traffic even if the INTERFACE directive is placed in the smb.conf file. This means it is possible to drive inbound traffic on both interfaces and have outbound traffic occurring only on the first interface.

To balance the load and have the appropriate inbound and outbound traffic on the correct interface, the two OSA cards must be connected to separate subnets. Using this configuration and the INTERFACE directive in the smb.conf file, a balanced load on the two connections can be achieved.

Scenario 1: Samba3 improvements for file serving

Objectives

The goal of these tests was to verify the viability of the IBM zSeries as a platform for large scale server consolidation. New installations may benefit from these methodologies to properly size and implement their new environment. The primary workload used in this project was defined by an IBM internal workload generator. The nature of the tool is to supply a varying number of users that access multiple file sizes and drive OSA and CPU utilization.

The testing was to answer these questions:

1. Did the performance improve when using Samba 3 in both a file serving and print serving environment?
2. What was the memory consumed by idle users?
3. Did the enhancements made to Samba alleviate the resource constraints reported by the user community?

The test environment was configured to stress the Samba server and may not be representative of a normal customer workload.

The throughput was as expected for Samba V3; see Figure 5. As the file size increased, the throughput rate as measured by completed file request transactions/sec dropped because there are more bytes of data to process for each file. On the small files, the overhead of the initial open and close overshadowed the actual data transfer.

Summary of test results

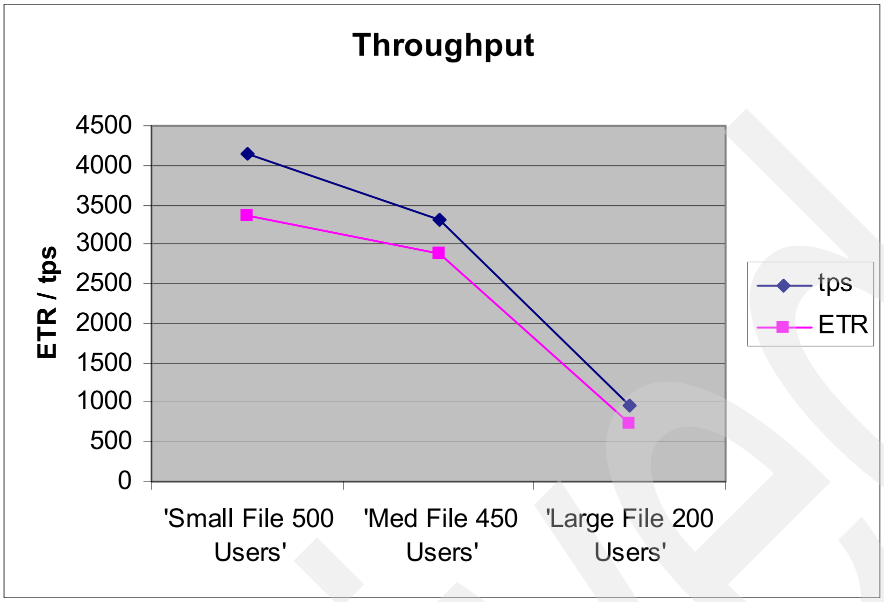

Figure 5 Throughput rate for small, medium and large files

The handling of large files was more efficient; see Figure 6. Larger files lead to higher bandwidth (MB/sec) of transfer rate. It takes fewer CPU cycles to process each byte of a larger file because of the fixed overhead to open and close the files, regardless of size.

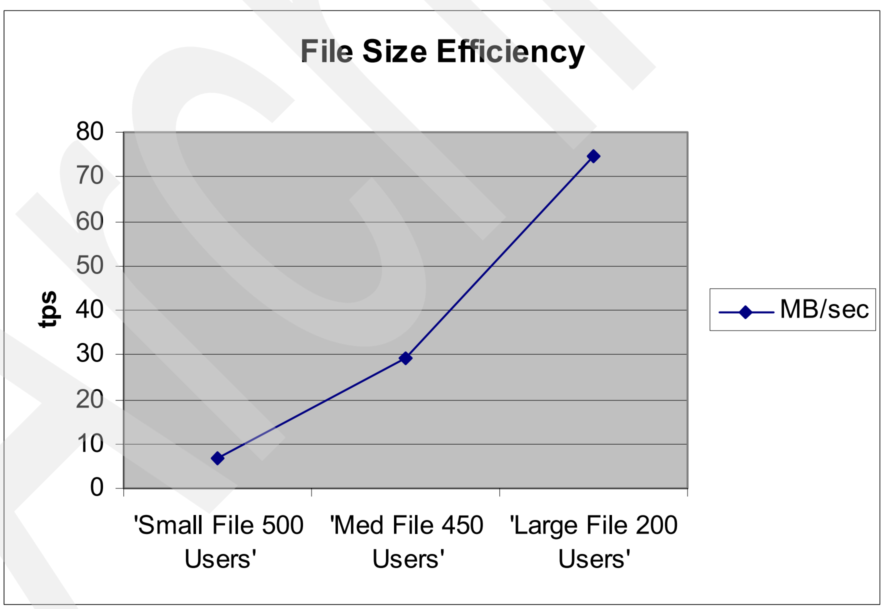

Figure 6 Efficiency of large file sizes

Enhancements were made to Samba V3 to reduce the memory consumption by idle users. Each Samba Version 3.0.5 idle user consumed approximately 390k. Tests show the memory requirements scale linearly as the number of users increase; see Figure 7.

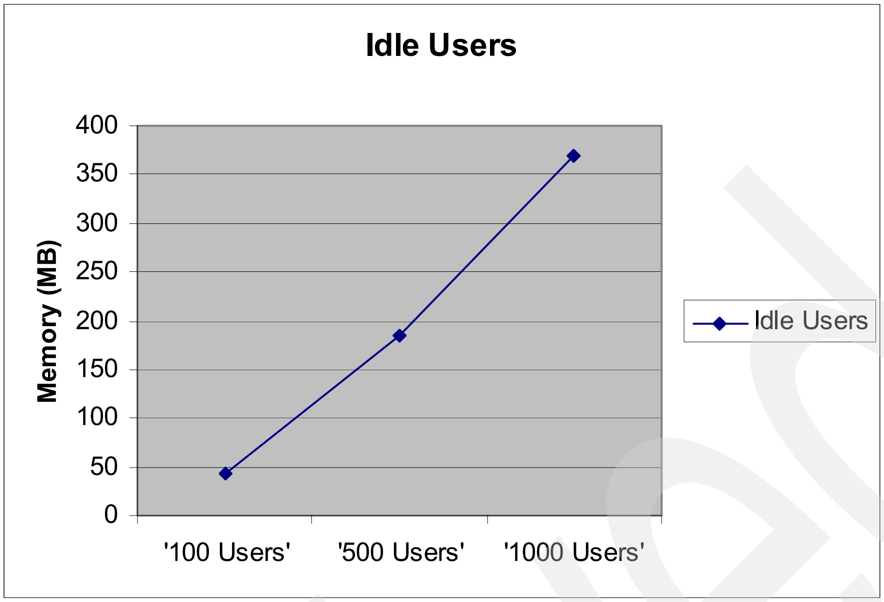

Figure 7 Memory use by idle users

Scenario 2: Print serving

Print serving was tested by sending the print jobs to a "virtual printer" with the data written to /dev/null. The virtual printer was another Samba server and CUPS running on an xSeries system. This workstation was placed on a 10/100 Ethernet LAN to simulate the majority of customer environments. The print server was on a Gigabit LAN.

Using multiple processes to submit print requests, requests would be queued on the server and the "virtual printer" would lag behind completing the prints. The 10/100 LAN nearly exhausted 100% of its bandwidth. Starting with ten simultaneous processes, the gating factor is the printer and it was still completing requests fifteen minutes after the workload generator had ended.

Conclusions and recommendations

Our tests indicate the Samba V3 application running on a zSeries host is viable for file and print serving environments. The tests showed significant performance improvement in throughput from Version 2.2.8. Idle Samba version 3.0.5 users did not consume excessive memory, and memory usage scaled linearly as users were added.

The Print Serving tests showed the throughput was always gated by the printer. The Samba V3 server can drive volumes that will fully utilize a 10/100 Ethernet LAN.

Our optimal configuration included:

- 1 Gb of memory
- SUSE SLES 8 Linux
- Linux kernel upgraded to 2.4.21-241
- Samba Version 3.0.5 with a log level of 0
- LVM with a 64KB stripe size
- ext3 file system, default journal parameters

- tuning parameters as documented in the smb.conf file
- a Gigabit Ethernet OSA interface

The authors of this Redpaper

Nin Lei is an Architect at the IBM zSeries Benchmark Center, conducting customer studies. He drives performance analysis work for all types of projects, and provides architectural guidance to customers in designing their applications to meet high performance and scalability requirements. He also advises customers on new technology exploitation issues.

Duane Mauler is a Certified I/T Specialist working to promote Linux on the zSeries platform. He has been in I/T for over 20 years, working across many different computer platforms.

J B Mills has worked for IBM for 27 years. His current specialty is zSeries Linux architecture. He has been involved with Linux on the mainframe since IBM announced its involvement in Open Systems on the Mainframe. He has experience in Samba installation and tuning, as well as z/VM® and WebSphere® tuning.

John Mullin has been a Staff programmer with IBM for six years. He is experienced in custom end-to-end environment designs and setup for client-based simulations.

Gene Ong is Linux Professional Institute Certified Level One. He works in the zSeries Benchmark Center at IBM in Poughkeepsie, NY

Judy Viccica is a Certified zSeries I/T Specialist who has been with IBM for 22 years. Currently she works in the zSeries Benchmark Center in Poughkeepsie, providing pre-sales technical support for client benchmarks and proofs of concept.

Special thanks to the following people for their contributions:

Donald Corbett, Senior Software Engineer
IBM zSeries Performance Analysis and Design

Michael Grundy
IBM Advanced Linux Response Team (ALRT)

Ann Jackson
IBM zSeries WW Technical Support - Linux

James Kuchler, tester
IBM Performance for zVM/Linux Workloads

Michael MacIsaac
IBM Linux on zSeries marketing technical support

Darren Swank, PMP®
IBM zSeries Benchmark Center

Appendix A

smb.conf is the main Samba configuration file. You can find a full, commented version at /usr/share/doc/packages/samba/examples/smb.conf.SuSE.

```
# Date: 2003-10-27
[global]
   workgroup = Client
#  security = DOMAIN
   security = PUBLIC
#     password server = Client20
   netbios name = TEST1
   os level = 2
   log level = 0
   log file = /etc/samba/log
   interfaces = xxx.xxx.xxx.xxx/255.255.255.0 xxx.xxx.xxx.xxx/255.255.255.0
   time server = Yes
   unix extensions = Yes
   encrypt passwords = yes
   map to guest = Bad User
   printing = CUPS
   printcap name = CUPS
   #socket options = SO_KEEPALIVE IPTOS_LOWDELAY TCP_NODELAY
   socket options = SO_KEEPALIVE IPTOS_LOWDELAY TCP_NODELAY SO_SNDBUF=14596 SO_RCVBUF=14596

   max xmit = 65536
   wins support = No
        #winbind uid = 10000-20000
        #winbind gid = 10000-20000
        #winbind separator = /
        #winbind use default domain = yes
   veto files = /*.eml/*.nws/riched20.dll/*.{*}/
[homes]
   comment = Home Directories
   valid users = %S
   browseable = No
   read only = No
   create mask = 0640
   directory mask = 0750
[printers]
   comment = All Printers
   path = /var/tmp
   printable = Yes
   create mask = 0600
   browseable = No
[print$]
   comment = Printer Drivers
   path = /var/lib/samba/drivers
   write list = @ntadmin root
   force group = ntadmin
   create mask = 0664
   directory mask = 0775
[aprinter]
   path=/aprinter
   printable = yes
   writeable = yes
   print command = "/bin/recprint %s"
[ashare]
   path = /ashare
   writable = yes
```

```
   guest ok = yes
[bshare]
   path = /bshare
   writable = yes
[cshare]
   path = /cshare
   writable = yes
```

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of 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.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Send us your comments in one of the following ways:

- Use the online **Contact us** review redbook form found at:
  ibm.com/redbooks
- Send your comments in an email to:
  redbook@us.ibm.com
- Mail your comments to:

IBM Corporation, International Technical Support Organization
Dept. HYJ Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400 U.S.A.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

eserver®	iIBM®	xSeries®
eserver®	Redbooks (logo)	z/VM®
FICON®	WebSphere®	zSeries®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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