Performance Analysis of Lotus Domino Server on Microsoft Windows 2000

By Clement Moy

Executive Summary

To help customers select and configure an appropriate server for Lotus Domino™ Server deployment, IBM publishes a Configuration and Options Guide and a sizing guide for IBM eserver x-Series. The configuration guide provides a recommended user capacity along with a range of required system configuration to accommodate variety in the application environment. It is updated for every server or family of servers at release of the products, giving customers server selection guidance. The sizing guide for Domino Server R5 prompts the user to quantify the job mix of the intended installation and then recommends one or two systems suitably configured. The sizing tool's purpose is to enable customers to make initial estimates before they engage an IBM sales team for a more in-depth analysis of their needs.

Although these tools provide timely and easy-to-understand information, they have limitations. The information is either static, as in the case of the configuration guide, or it makes no provision for the customer to pre-select the server model. Many customers prefer to standardize their installation on a specific server model that they have spent time and resource to certify for their business. What many of them need is appropriate information or technical consultation that helps them configure their choice of server model to meet their Domino Server requirements.

The purpose of this paper is to provide appropriate data and guidelines that I/T specialists and IBM presales technical support specialists can use to tailor sizing recommendations to their customers' requirements. This paper provides measurement-based performance information that can address some of the most common questions about capacity planning in a Lotus Notes® environment.

Information presented is intended as a performance guide for trained technical support personnel and should be used with consideration for the production environment for which the sizing is intended. It is assumed that the reader has a sound knowledge of Intel®-based servers and major subsystems, operating systems, performance concerns of server subsystems, and Lotus Domino Mail Server. Readers who need a review of server performance tuning on Microsoft® Windows NT® and Windows® 2000 are urged to read the IBM redbook *Tuning Netfinityâ Servers for Performance - Getting the Most Out of Windows 2000 and Windows NT 4.0 (* SG24-5287-01). Successful sizing ultimately depends on the reader's skills and knowledge of the customer's environment. Technical information for Intel-based servers is available on IBM's Web site for xSeries servers.

All measurements were made using the Domino Server R5 Mail workload. Analysis of the measurement results is applicable only to the Notes mail environment. For sizing of Domino Server used to support applications other than Notes Mail (e.g., WebMail , Database), the reader will also need system capacity information for other applications that can run on a Domino Server. This type of information is beyond the scope of this document, but it is regularly published by Lotus Development Corporation on their Notes.net Web site.

September 2001

Introduction

The workload used for this analysis is similar to that used for benchmarking on Domino Mail Server Release 5. Some of the parameters used in the measurements for this report represent a more recent specification of a typical mail customer environment.

It is important to note that, according to Lotus, Domino Server R5 supports about 10,000 users per single partition. Published results confirm this; for example, the IBM Netfinity 7600 server achieved 10,045 users on a single Domino partition on Windows 2000 Advanced Server. Support of more than 10,000 users on a single physical server requires deployment of multiple Domino Partitioned Servers. Sales and field technical support teams, to date, have relayed little customer interest in using Domino Partitioned Server as a solution to scaling up a Notes mail installation. An industry survey commissioned by Lotus showed that typical Domino customers deploy no more than about 2,500 registered users per Domino Server partition. IBM's iSeries, pSeries and zSeries servers can support multiple such partitions within one physical server.

However, with the xSeries on the Windows platform, the single partition usually translates into one server. For this reason, the measurements described in this report were taken on a server running a single Domino Server partition for user counts no higher than 10,000, thus minimizing the possibility of the software becoming a performance bottleneck.

IBM Advanced Technical Support specialists have been recommending uni-processor and two-way SMP servers for their large Domino Server installations. They recommended four-way Intel® Pentium® III Xeon™-based servers with two to three 550MHz CPUs before the introduction of the 1GHz and faster Pentium III-based two-way servers, which can easily support the same user load. A four-way SMP server with the 700MHz Pentium III Xeon processor and a two-way SMP server with the 1GHz Pentium III processor were employed for different groups of measurements in this study.

Following are some of the parameters and practices implemented for this performance study:

1. Domino Server services started: Router Task
2. Performance metric used for comparison: maximum number of users supported for a response time under 5 seconds
3. Mail message size: 25 Kbytes (vs. the 10 Kbytes normally used in benchmarking)
4. Disk subsystem analysis restricted to the disk volume holding user mail
5. 25 to 70 percent of the disk volume was populated with actively used data
6. Logical disk Allocation Unit Size: 16 Kbytes

See the Appendix for additional details.

Benchmark Environment

Server Configurations

Table 1

	System A	System B
Model	xSeries 350	xSeries 340
Processor/Cache	4 x 700 MHz Pentium III Xeon / 2MB L2	2 x 1GHz Pentium III / 256KB L2
Memory/Speed	4GB / 100MHz	4GB / 133MHz
Network adapter	2 x 10/100 Mbps Ethernet Adapter	2 x 10/100 Mbps Ethernet Adapter
Boot volume (Windows 2000 Advanced Server, Domino code and main data directory)	RAID-1 on ServeRAID™-4L	RAID-1 on ServeRAID-3L
Data volume (user mail files)	2 x ServeRAID-4M, 4 x EXP300 enclosure, each holding 14 x 15Krpm drives	Memory Scaling: 2 x ServeRAID-4M, 4 x EXP300 enclosure, 54 x 15Krpm drives Disk Scaling: 1 x ServeRAID-4M, 1 x EXP300 enclosure, up to 14 x 15Krpm drives

We projected that a two-way SMP server with a 1GHz processor, configured like System B, should be able to support 10,000 Domino R5 Mail users. In our measurements, System B supported only 9,850 users. As expected, System A, with only two 700MHz processors installed, supported 9,700 users. We used System A for the processor scaling analysis so that we could have a maximum user count of 10,000 in the comparison data for processor scaling and memory scaling.

With either system, the server under test (SUT) was configured with a boot volume on a RAID-1 array defined on an IBM ServeRAID-3L or IBM ServeRAID-4L controller. Since the Average Disk Queue Length of the boot volume during these measurements was typically less than 0.1, any difference in server performance due to the difference in the RAID adapter used for the boot volume is negligible. In addition, the disk I/O activity from the boot volume accounted for only 1 to 2 percent of the total disk I/O activity in the system. The disk subsystem consisted of the following:

1. Logical drive C:, defined on a RAID-1 array that held the operating system, the Domino Server code and main data directory, containing the directory and log file.
2. Logical drive D: for disk subsystem analysis. Logical drives D:, E:, F: and G: for processor and memory scaling analysis. With a couple of obvious exceptions, all drives used were 15K rpm 18.2GB drives.

Except for the smallest Domino Server configuration, partitioning the disk subsystem in this manner is often recommended because it facilitates system performance optimization. In this performance study, it allowed us to isolate the server activity that best illustrates the performance effect of different disk subsystem configurations.

System configuration used for evaluating processor and memory subsystems

System A was used for all processor and memory scaling measurements. The mail data storage consisted of four arrays configured as logical volumes D:, E:, F: and G:. Mail files were evenly distributed across the four arrays connected to the server through two ServeRAID-4M controllers.

For these measurements, the resulting average queue length per physical disk was consistently below 1, ensuring that the disk subsystem was never a performance bottleneck.

Clients were labeled either odd or even according to the order in which they were started and distributed over two 100Mbps LAN segments. The SUT was connected to both client LANs. A controller client and additional support servers were located on a third LAN segment. All three LAN segments were connected to a Windows 2000 Domain Controller/Domain Name Server that also served as a router during test setup.

During a 10,000-user run, the combined data rate of both Ethernet cards on the Domino server averaged 17Mbps with a peak of 20Mbps, far from saturating a single 100Mbps full-duplex Ethernet LAN segment, much less the two LAN segments used for the measurements. As such, the network was not a performance bottleneck.

System configuration used for evaluating disk subsystems

Measurements for this analysis were made on System B because faster processors with speeds 1GHz and higher are more likely to be encountered by the reader in server sizing exercises. Processor power and memory capacity were monitored to ensure that they were never a performance bottleneck.

The disk subsystem was partitioned as is frequently recommended and commonly practiced. It consists of a boot volume (a small array holding the operating system, application code and common files), and a data volume (a separate array holding the mail files). The boot volume was under-utilized by the workload whereas the data volume accounted for 98 to 99 percent of the total disk activity. Evaluation of different disk subsystem configurations was performed only on the data volume. The size of the array used as the data volume was kept small enough so that its maximum performance capacity did not exceed the limitations of the application. Otherwise, the software limitations can invalidate performance comparisons.

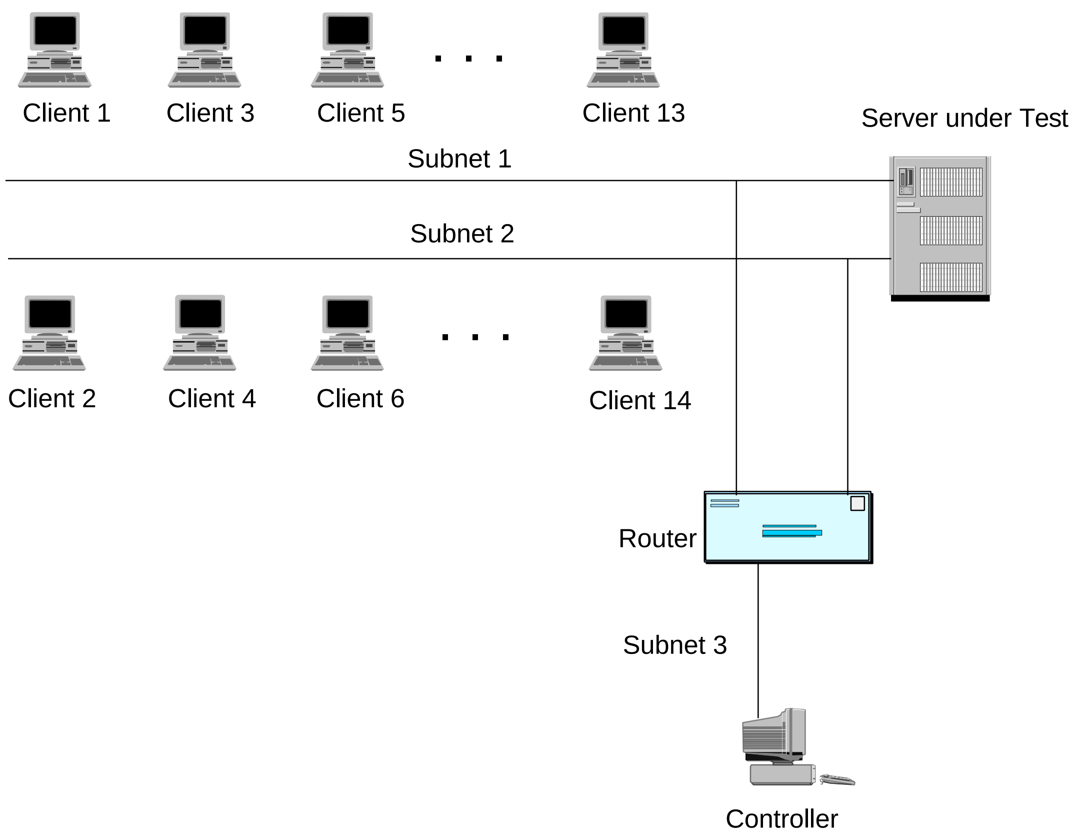
All measurements were made using 15K rpm 18.2GB Ultra160 SCSI drives, except when the mail file volume was built on a two-drive RAID-1 array. To hold a sufficient number of files to saturate a data volume built as a RAID-1 array, we needed drives larger than 18.2GB. Since 15K rpm 36.4GB drives were not available at the time, we substituted 10K rpm 36.4GB drives. Then, we measured the performance capacity of both 10K rpm drives and 15K rpm drives on 14-drive RAID-1E arrays. Finally, we used the ratio from these two measurements to adjust the performance capacity of 10K rpm RAID-1 array accordingly.

LAN activity was not a performance bottleneck in System B.

Network performance scaling

We did not perform a scaling study on the network because the Notes Mail workload did not produce a sufficient amount of activity on the LAN to show visible effects of changes in network subsystem configuration. Under the heaviest user load in our study (10,000 users), the two network adapters used managed only a combined average data rate of 17Mbps (peaking around 20Mbps) across both Ethernet LAN segments.

Measurement testbed



Workload

Other than specifying a 25-Kbyte message size, the workload is basically the same as that used in Domino R5 Mail benchmarking. Substantial file growth occurred during the runs. For example, in an 8,800-user run using a 14-drive RAID-1E array, 9,000 mail files each started at 7MB initial file size grew an average of 54 percent during a 10-hour run. A detailed description is provided in the Appendix.

Considerations for using the measurement information

In using the measurement data to size a configuration, it is important to understand the customer's environment and to exercise judgment accordingly. Following are some guidelines for using the measurement data presented in this paper:

1. The Domino Server used in this study supported only locally connected Notes clients using the Note Remote Procedure Call (NRPC) protocol. Many Domino Mail servers may be supporting mobile users, vendors, business partners and clients all within the same system. These clients, such as Microsoft Outlook, Netscape Communicator and Internet Explorer, use SMTP/POP3, IMAP or HTTP protocol. Based on Lotus' own measurements, most of these other mail protocols are more resource-intensive than the NRPC protocol. The article, "Optimizing Server Performance, I/O Subsystems," published August 2, 1999 on Notes.net, showed that on unconstrained systems, the maximum achievable number of mail users for different mail protocols is 10,000 for NRPC; 2,000 for

HTTP; 3,000 for IMAP; and 5,000 for SMTP/POP3. Sizing a Domino Mail server for a mix of these users is not directly supported by the information provided in this document.

2. The performance metric used in this analysis is the maximum number of users supported at an acceptable response time. As will be quantified later, for any new subsystem configuration, we aim for a measurement error rate under 2 percent, but the actual measurement could have an error rate as high as 4 percent. Performance capacity of a server is shown as multiples of 100 users. In rare cases, where the total user count is low, a multiple of 50 users was used.
3. Domino Server experiences a surge of activity when a user logs in and opens his mail file. All the other activities in the benchmark after that initial phase impose 15 to 20 percent less load. The rate at which users are applied affected the maximum numbers of connected users achieved during these measurements. For consistency in measurement criteria, we controlled the rate of user login to maximize the users supported by the system during steady state. The reader needs to discount the performance capacity shown if fast user login in the customer deployment is expected.
4. Throughout this document, the author shows relationships as ratios of the measurements obtained with these tools. The resolution of the ratios shown does not imply a corresponding level of accuracy. This point will be reinforced many times throughout this document.
5. When comparing two measurements, the 2 to 4 percent uncertainty in each data point compounds to a propagated error of 2 to 8 percent in the comparison. The example in Table 2 illustrates a worst-case scenario.

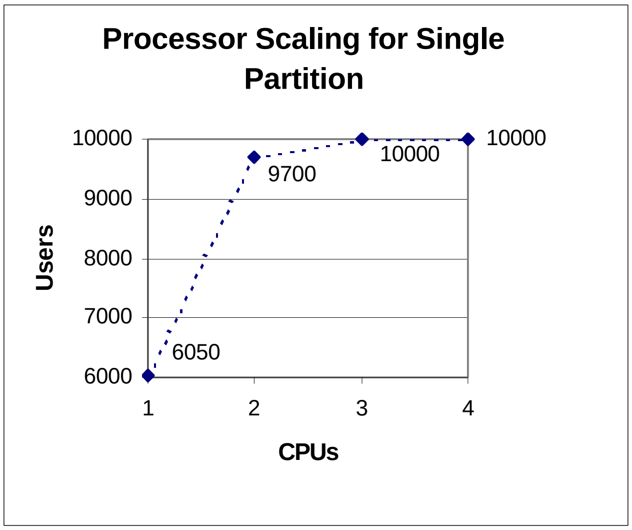
Table 2

Data Point A	Data Point B	Ratio A/B	Data Point A Error	Data Point B Error	Ratio A/B
1	1	1	+4%	-4%	1.08

Results and Analysis

Processor scaling

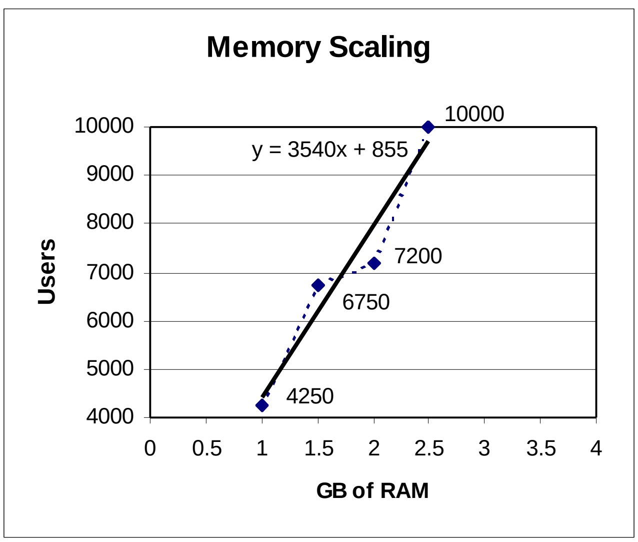
For all measurements, there were 10,000 registered users in the directory and 10,000 mail files in the mail file storage. The 10,000-user design limitation of Domino Server R5 was reached with three processors in System A, at an average processor utilization of 67 percent. Maintaining the same user load and adding the fourth processor lowered the average processor utilization to 52 percent for all four processors. Intuitively, total processor utilization is equal to the product of the number of processors used and the average processor utilization. In terms of number of processors used, total processor utilization was 201 percent (3 x 67 percent) when three processors were used and 208 percent (4 x 52 percent) when four processors were used. This helps explains why three processors were sufficient to support 10,000 users, with slightly more processor resource consumed due to the resource management associated with one more processor.



Memory Scaling

System B was used for these measurements. Four 14-drive RAID-1E arrays were connected to two ServeRAID-4M controllers for mail data. Selected measurements were verified using the four-way xSeries 350 with 700MHz Pentium III Xeon processors.

The following graph shows the maximum number of users supported at memory sizes of 1GB, 1.5GB, 2GB and 2.5GB.



The limit of the application's ability to use memory was reached at or below 2.5GB. At 10,000 users, 2.5GB of memory was enough to keep average user response time to 109 milliseconds. Increasing memory from 2.5GB to 4GB while keeping the user count at 10,000 improved average user response time by an insignificant amount. A solid trend-line is added to show that the scaling is approximately linear up to the software limitation in the application.

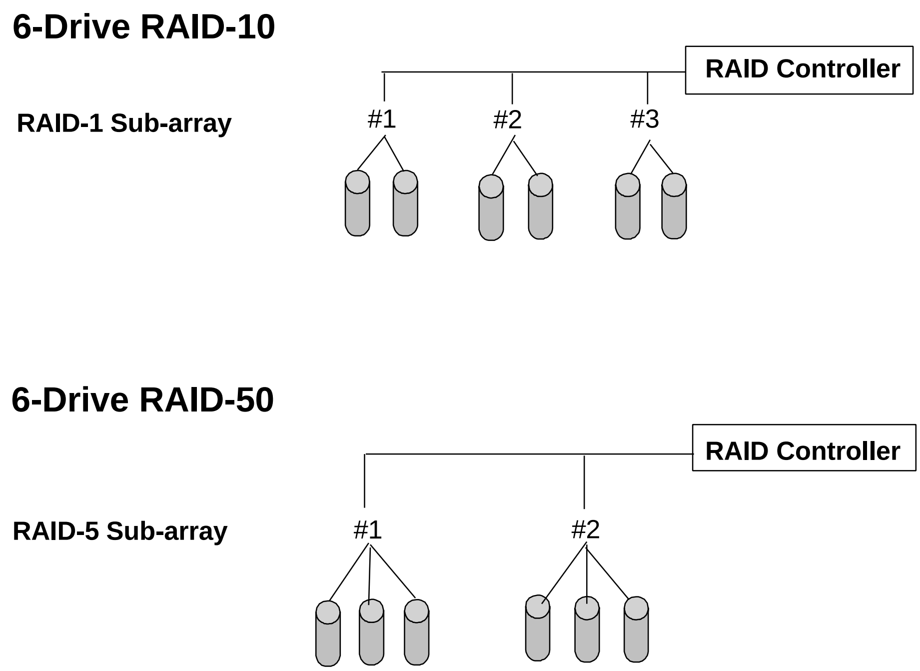
Domino Server R5 does not take advantage of the 3GB application memory feature of Windows 2000 Server, or the 8GB physical memory support in Windows 2000 Advanced Server. By design, it can only take advantage of up to 2GB of application memory in Windows. Additional memory beyond 2GB serves only to provide dedicated memory for Windows and the system cache. Because Domino Server manages its own data cache in its NSF buffer, which uses memory within the 2GB application memory, it receives limited benefit from a large system cache. In Domino Server benchmarking under Windows NT and Windows 2000 Server, we consistently see no performance benefit in using more than 2.5GB of physical memory when a single Domino Server partition is employed, since the operating system used 512MB and Domino used 2GB.

Disk subsystem performance analysis

For these measurements, we stressed the disk subsystem heavily so that performance variations due to changes in the disk subsystem configuration were maximized. In addition, we kept the total number of users within the range supported by a single Domino Server partition. With these constraints in mind, we used the maximum number of drives that can be configured in a 14-bay IBM EXP300 Storage Expansion Enclosure. For RAID-1E analysis, we used a maximum of 14 drives. For RAID-5 analysis, we used a maximum of 12 drives because a 12-drive RAID-50 is the largest RAID-50 with 3-drive sub-arrays that can be configured in the EXP300. As will be shown later, we found that there was no performance advantage in splitting the backplane of the EXP300 enclosure. The simple configuration of one array on one SCSI channel was used for these measurements.

The reader who wishes to consider any one of the array configurations analyzed here, but needs to support more users, should consider using multiple such disk arrays.

To illustrate the configuration of RAID-10 and RAID-50 built using 3-drive RAID-5 sub-arrays, the following diagrams show six drives configured as RAID-10 and RAID-50 with 3-drive sub-arrays.



Conceptual diagram of RAID-10 and RAID-50 (spanning 3-drive RAID-5)

Standard physical disk configuration for mail data

For RAID-1 and RAID-1E analysis, up to fourteen 15K rpm disk drives were connected to the RAID controller through one SCSI channel. For RAID-5 and RAID-50 analysis, up to twelve 15K rpm disk drives were used. Since this workload has an average transfer size of about 12 Kbytes, a 16-Kbyte stripe-size was used to minimize the average number of disk transfers per I/O access. The controller cache was set for write-back, which has consistently shown better performance than write-through mode for this workload. The performance difference was negligible whether the mail data volume was formatted using a 512-byte or 16-Kbyte allocation unit size (specified when formating the disk by command line as FORMAT D:\ /ALLOC: 16K). For all measurements conducted in this study, the mail data volume was formatted with a 16-Kbyte allocation unit size.

Run-to-run variability

Using the standard disk configuration for RAID-1E and a 50-Kbyte Notes mail message size, two runs with identical number of users were conducted. A change in throughput (completed transactions per second) of less than 1 percent was observed between the runs.

Measurement accuracy

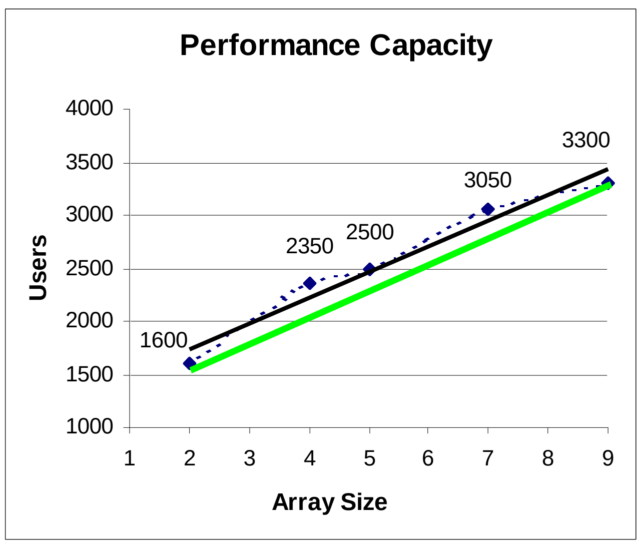
While the system throughput measurement (i.e., completed transactions per second), can be reproduced with about a 1-percent accuracy if the number of users is kept constant, determining the maximum number of users that can be supported by any one system configuration with the available tools is a very time-consuming process involving educated guessing of the outcome and a fair amount of trial and error. The capability of the tool and the process used has a 2-to-4 percent inaccuracy in determining the maximum performance capacity for each server configuration. Achieving greater accuracy involves a disproportional increase in the time required to conduct each measurement and for the purpose of these studies was not justified.

Effect of array size on performance

Justifying the estimation methodology

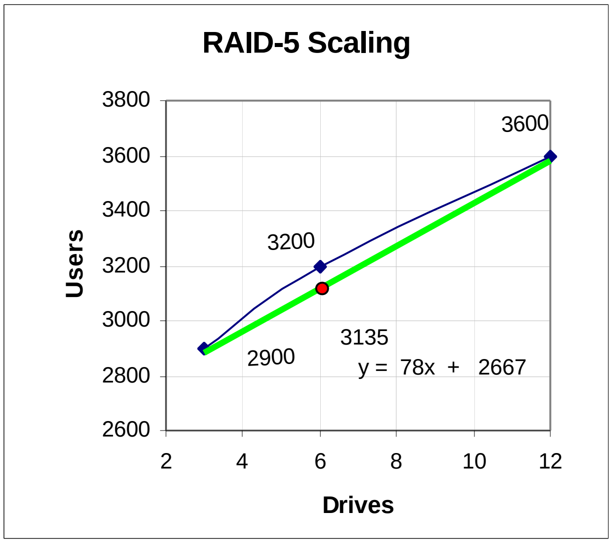
From performance capacity measurements made for RAID-1E arrays of 10K rpm drives, we observed that the number of users supported by an array increases linearly for a low number of drives and logarithmically as the array size increases beyond 9 to 10 disk drives. However, detailed analysis shows that estimates using a linear trend-line (as provided by the spreadsheet used to generate the following chart) produce acceptable accuracy. In the following graph, the linear trend-line is indicated by the black line drawn through the measurements. If only the measurements at both ends (array size equal to 2 and 9) are available, the linear trend-line will simply be the thick lighter straight line joining the two end-points.

While a linear approximation based on only two measurements provides a less accurate estimate than one based on additional intermediate measurements, the reader can see that two lines are roughly parallel to each other, meaning that the rougher approximation (lighter line) will consistently produce estimates that maintain the same accuracy in relation to the estimate based on more data measurements (dark line). As we will demonstrate with the actual measurements from this study, a linear interpolation with only a data point at each end will frequently be adequate for estimating the intermediate data points.



Performance capacity of RAID arrays

The following graph illustrates the estimation of performance capacity for a 6-drive array from measurements made for a 3-drive array and a 12-drive array. Assuming we only have measurements for the 3-drive array and the 12-drive array, the equation for the linear trend-line (thick straight line) is shown in the graph. Using this equation, and the value 6 as “x,” the estimated performance capacity for a 6-drive RAID-5 array would have been 3,135 users. Since the measured performance capacity for a 6-drive RAID-5 array is 3,200 users, this estimation method resulted in a 2-percent error.



For comparison, the measured performance capacity for RAID-5, RAID-50, RAID-1E and RAID-10 arrays of selected sizes are shown together in Table 3. Array sizes that are invalid for specific RAID schemes are indicated as N/A for "not applicable." Using the 18.2GB drives, accurate measurements were not possible for RAID-1, RAID-1E and RAID-10 arrays that contained eight drives or less because there was insufficient storage space to determine the array's performance capacity. In such cases, the user count reached before the server ran out of mail file storage is shown with a "+" sign next to the performance capacity value. The average user response time at such user counts was too low for the disk array to be saturated.

The disk subsystem was generally heavily saturated for measurements made on 14-drive RAID-1E and RAID-10. The effect is that the measurement errors inherent in our procedure often overwhelmed the performance difference between a 12-drive array and the corresponding 14-drive equivalent, preventing a meaningful comparison between 12-drive 14-drive RAID. For this reason, measurements from 14-drive RAID are not shown.

Table 3

Performance Capacity	2 Drives	3 Drives	6 Drives	8 Drives	12 Drives
RAID-5	N/A	2,900	3,200		3,600
RAID-50 (3-drive sub-arrays)	N/A	N/A	4,000	N/A	5,200
RAID-1/RAID-1E	3,118[1]		4,300+		7,700
RAID-10	N/A	N/A		7100+	8,700

[1] Measured using 36.4GB 10K rpm drives and adjusted to provide an estimate for 15K rpm drives.
The "+" sign Indicates that actual performance capacity could be higher if higher-density disk drives were used.

For the array sizes shown, estimation of performance capacity for array sizes not measured generally produces an answer within the margin of error of the measurement method. For example, using measurements for a 3-drive RAID-5 and a 12-drive RAID-5 to estimate the user count for a 6-drive RAID-5 yields 3,133 users, about 2 percent less than the measured 3,200 users

If two measurements used for linear interpolation of a data point in between them are closer together on the x-axis, the estimation is more accurate. However, the difference is not significant. For example, if the performance capacity for the 9-drive RAID-5 array was first estimated using measurements for the 3-drive RAID-5 and the 12-drive RAID-5, and then estimated using the measurements for the 6-drive RAID-5 and the 12-drive RAID-5, the differences between the two estimates, 3,367 and 3,400, respectively, are only about 1 percent of each other.

Implication for sizing

RAID-1E vs. RAID-5 need not be a sacrifice of cost for performance.

While RAID-5 has been commonly regarded as the RAID scheme with the best price/performance, we have shown such large performance gains over RAID-5 with RAID-1E that it could more than compensate for the lower percentage of usable storage in a RAID-1E scheme. When you take into account relationship between the price of disk drives of different capacity, RAID-1E can be a more attractive alternative.

As an illustration, notice in Table 3 that a 6-drive RAID-1E array supported more than 4,300 users whereas it took a 12-drive RAID-5 array to support 3,600 users. Aside from the lower number of drive units, there is an additional cost savings realized with the 6-drive RAID-1E due to lower storage enclosure cost, but what about the reduced usable storage of RAID-1E when compared to RAID-5?

Suppose, for example, that an installation needed the amount of storage provided by a 12-drive RAID-5 array built with 18.2GB drives in order to support 3,600 mail users. We have already established that a 12-drive RAID-5 array built with 18.2GB 15K rpm Ultra160 SCSI drives satisfied the system capacity requirement. We have also determined that a 6-drive RAID-1E array built with the same drives will support more than 3,600 mail users while providing more than half the required amount of storage space. Substituting 36.4GB 15K rpm drives for the 6-drive RAID-1E solution will more than satisfy the storage space requirement.

RAID schemes involving spanning

RAID-10 has the same advantage as RAID-1E over RAID-5. In addition, it provides a bonus of higher performance. The decision to use RAID-10 should be simple if supported by the RAID controller. RAID-50 provides a solution midway between RAID-5 and RAID-1E in usable-to-purchased storage ratio, and performance. Its applicability is left to the imagination of the reader.

Comparison of RAID schemes

The relative merit of RAID-50 using 3-drive RAID-5 sub-arrays, RAID-1E and RAID-10 over a RAID-5 array of the same size is shown in Table 4. For each array size, the increase in user capacity achieved by each RAID scheme compared to RAID-5 is shown in Table 4. A 6-drive RAID-50 array, using 3-drive RAID-5 sub-arrays, supports 25 percent more users than a 6-drive RAID-5 array. A 6-drive RAID-1E supports 44 percent more users, and so on.

Table 4

Performance Gain over RAID-5	6 Drives	12 Drives
RAID-5	0	0
RAID-50 (3-drive sub-arrays)	25%	44%
RAID-1E	40%	114%
RAID-10	-	142%

In measurements for a 12-drive RAID-10 and RAID-1E, RAID-10 showed a 13-percent performance advantage over RAID-1E.

Sensitivity to allocation unit size

An array-configuration option that can affect disk subsystem performance is the allocation unit size used to format the logical volume defined on the array. We measured the performance capacity of 14-drive RAID-1E array and 12-drive RAID-5 array using allocation size of 16 Kbytes and 32 Kbytes, respectively. There was no measurable difference for RAID-1E arrays and a difference of only 2 percent for a RAID-5 array, within the measurement margin of error.

Sensitivity to message size

While measurements were made using a constant message size of 25 Kbytes, we wanted to show readers the effect of doubling the message size so that they can adjust the information we provide for a deployment known to use a different average message size. We found that a 14-drive RAID-1E array supported the same number of maximum users when the message size was changed from 25 Kbytes to 50 Kbytes. At the same time, a 12-drive RAID-5 array supported about 8 percent fewer users when the message size was changed similarly.

Table 5

RAID Scheme	Change in number of users when message size changed from 25KB to 50KB
14-drive RAID-1E	0%
12-drive RAID-5	- 8%

Benefit of two SCSI buses over one

Like most client/server applications, Notes Mail is an application that usually does not apply a heavy load to the SCSI bus. Our measurements showed that the 14-drive RAID-1E was within 7 to 8 percent of its performance capacity at 7,200 users. At this level of system load, if the RAID controller has an unused SCSI channel, as we had with a ServeRAID-4M controller, it represents an attractive possibility for performance improvement by distributing the existing disk drives over an additional SCSI bus.

For comparison using RAID-1E, two runs were made with 7,200 users, using a ServeRAID-4M controller. The first run was made with all 14 drives connected to the same RAID controller channel. For the second run, the 14 drives were evenly divided across two channels.

The two-channel configuration supported 7,300 users (vs. 7,200 users for the single-channel configuration) while maintaining approximately the same average user response time. This was only a 1.4-percent increase. Considering the 2-to-4 percent inaccuracy in our measurement process, this is not a noticeable performance difference. When we kept the number of users the same, the two-channel configurations showed a 73-percent drop in average user response time. While the 73-percent drop in average user response time by itself seems significant, it only translated to a 1.4-percent spare capacity.

In the two-channel configuration, at 7,300 users, the array holding the mail files conducted an average of 1,021 disk transfers per second at an average size of 11,580 bytes per transfer. This translates to a SCSI bandwidth requirement of 11.3MB per second. When all the drives were connected to the controller through the same SCSI channel, the bandwidth requirement was only around 22.6MB per second, far from saturating the 160MB/sec peak throughput of the SCSI channel on the ServeRAID-4M controller. Low SCSI bus utilization diminished the performance benefit of additional SCSI bus bandwidth.

Potential improvement in customer satisfaction from an additional SCSI channel

Besides the significant drop in average user response time, we also noticed a large difference in the distribution of the average user response times. The swing in Maximum Average User Response Time reported for each minute by the clients was noticeably larger when only one SCSI channel was used. This was true for both RAID-1E and RAID-5 schemes. The maximum response time reported by a client is the maximum response time reported for the group of users simulated by that client.

At a time when return on investment is a major criterion in making I/T investments, it is unlikely that system performance will be ideal under all possible user loads. The author wishes to suggest that when average user response is slow, the users may be more annoyed by a significant difference in level of service (user response time) between users, or during different times of the day. To quantify this metric, variability in peak response time reported for each minute by each client is computed as the standard deviation of all minute-by-minute samples of peak response time measurements. The difference in using two SCSI channels vs. one SCSI channel is shown in Table 6 as a drop in the Standard Deviation. A smaller standard deviation means smaller fluctuation, which is desirable where Average User Response Time is concerned.

Table 6

Standard Deviation	1 Channel	2 Channels	Drop in Standard Deviation due to the Second Channel
RAID-1E	4,184	1,661	60%
RAID-5	661	24	96%

Write-back vs. Write-through

For both RAID-1E and RAID-5 schemes, switching the RAID controller cache from write-through to write-back increased the number of users supported by a small amount. This has been consistently observed during four years of Domino Mail Server benchmarking and at least one measurement in a customer production environment. On a 14-drive array, the measured difference between RAID-1E and RAID-5 is shown in Table 7.

Table 7

RAID Scheme	Gain in Users for Write-back Vs. Write-through
RAID-1E	2.9%
RAID-5	4.2%

15K rpm over 10K rpm

Many users will be substituting 10K rpm drives with 15K rpm drives with new server purchases. Thus, understanding the performance gain in terms of number of users supported as the result of a technology change is helpful in estimating disk subsystem requirements on system upgrades. We determined the effect of disk speed increase from 10K rpm to 15K rpm for 14-drive RAID-1E and 12-drive RAID-5 arrays. Table 8 shows the much greater performance gain under RAID-5.

Table 8

RAID Scheme	Percentage Gain in Users Supported When Upgrading from 10K to 15K rpm Drives
RAID-1E	7.5%
RAID-5	44%

Summary and Conclusions

With the current symmetric multiprocessor servers on the market, where processor speed is 900MHz or above, we have shown that a two-way SMP server is adequate for simple Notes Mail deployment using a single Domino Server partition at close to 10,000 concurrent users, the design limit for Notes mail in Domino Server R5. We have also shown that to support the maximum number of concurrent users, we need no more than 2.5GB of memory. We also noted that a typical Domino Mail Server deployment supports additional services and mail protocol, which generally use more system resources per user and significantly decrease the number of concurrent users supported.

Using the big difference in performance capacity of RAID-1E array over a RAID-5 array of the same number of drives, we have shown that a RAID-1E solution can provide much better price/performance. Furthermore, with array spanning as supported by the IBM ServeRAID-4 family of controllers, additional performance gains are possible with RAID-10 (over RAID-1E) with no loss to the percentage of usable storage space. We have also shown that RAID-10 and RAID-50 spanning 3-drive sub-arrays provide additional attractive alternatives that can enhance the storage solution.

We have examined various common disk subsystem configuration questions and quantified the effect of different decisions. In the case of the benefit of using two SCSI channels instead of one, we found a potential user satisfaction benefit even though the additional SCSI channel did not allow us to support significantly more users.

Finally, the reader should not miss the message that there is much uncertainty in the measurements and how well the workload used for these measurements represents their production environment. Following recommendations of the software developer and application specialist in making allowance for all the difference between your intended production environment and our measurements is crucial to making a sound system sizing.

Appendix

Domino Server services started

Lotus advises that idle Domino services have no noticeable performance cost other than for the small amount of memory they occupy. Our experience with Domino Mail performance analysis supports this; therefore, we started only the Router tasks from the administrator selectable task list for all measurements made.

Performance metric used for comparison

For maximum differentiation, we determined the maximum number of users supported by a server configuration for which the average response time, as reported for the users, is less than 5 seconds. We considered this the "performance capacity" of the particular server configuration. When the only subsystem being modified was the disk volume holding the mail files, we referred to the metric as the "performance capacity" of that disk volume or underlying RAID array. Between the methodology used and limitations of the tools, there is about a 2 to 4 percent inaccuracy in the determination of this metric.

Message size

In the SPECmail2001 benchmark specification released on January 24, 2001 by the Standard Performance Evaluation Corporation (SPEC), the user profile specified a weighted average message size of 25 Kbytes. Since this represents a recent user profile specification, we chose to use this message size in generating our mail workload. This message size is about 2-1/2 times as large as the message size used in benchmarking with Domino Server R5.

Disk storage configuration considerations

Since the disk activity from the mail file storage represented about 98 to 99 percent of the overall disk activity in the Domino Mail Server, we can obtain most the disk subsystem performance gain by focusing on the disk array or arrays used to hold the mail file. Having only one common file type within a disk array allows the system to derive maximum benefit from performance optimization on the array. For this reason, when evaluating performance effects of different disk subsystem configurations, we varied only the configuration of the mail file storage.

Population of disk storage

The mail file storage was populated to about 25 to 70 percent in the evaluation of different RAID configurations, depending on the RAID scheme and the number of disk drives used in the array. Mail files used by the active users were interspersed with the mail files of the inactive mail recipients to simulate the average access time of accessing the mail file of a random user among all registered users. Inactive mail recipients received mail from active mail users but did nothing else. All mail files start with a size of 7MB. Over a 10-hour run, during which 10,000 users were added in the first 8 hours, the average mail file grew about 54 percent.

Part of the reason for this level of disk population was our attempt to simulate a production environment that is sensitive to storage cost. This is in contrast to a benchmark environment where the active data often occupies 10 percent or less of the available storage. The other part of the reason was due to the maximum size of the disk drive available. In this analysis, the average seek-time of the disk drives is significantly higher than for a benchmark environment. A repeated observation is that the average response time sampled after ramp-up increased as the run progressed after fully ramp-up (usually 3 to 4 hours after it had been started). This is in contrast to the experience from benchmarking when average response time decreased after ramp-up. While this was true in a benchmarking environment, likely because of smaller mail message size and very low disk population in the data volume, it was not the case for this study.

Workload used for performance measurement

This Mail workload executes Notes transactions that model a mail server at sites that rely only on mail and calendar and scheduling for communication. It models an active user reading and sending mail, scheduling an appointment, and sending a Calendar & Scheduling (C&S) invitation. An average user will execute this script 4 times per hour. For each iteration of the script, there are 5 documents read, 2 documents updated, 2 documents deleted, 1 view scrolling operation, 1 database opened and closed, 1 view opened and closed, 3 messages comprised of 1 memo to 3 recipients, 3 lookups against the Domino Directory, 1 C&S appointment, and 1 C&S invitation are sent to the recipients approximately every 90 minutes. All documents, original or created, were uniformly 25,000 bytes long.

Mail recipient names were resolved to fully qualified Lotus Notes addresses using the server's copy of the address book instead of the user's local copy. Only connected users were active users, while all registered users were mail recipients. Their mail files received mail throughout the run.

References

1. "Optimizing Server Performance, I/O Subsystems," August 2, 1999, Notes.net.
2. SPECmail2001 Specification, Standard Performance Evaluation Corporation.
3. Technical information related to Intel-based servers, IBM Corporation.
4. Tuning Netfinity Servers for Performance - Getting the most out of Windows 2000 and Windows NT 4.0, IBM Redbook SG24-5287-01.

IBM Server Group

Department MX5

Research Triangle Park NC 27709

Produced in the USA

9-01