



Lynn F Behnke
Beverly Cusato
Dan Frissora
Dave Johnson
Joe Peterson
Walter Scanlan
Richard Smitherman

Sizing Large-Scale Domino Workloads on iSeries

In 2002 a large iSeries™ customer asked IBM® to recommend configurations for its expanding Domino® workloads. At the time we lacked performance data for environments similar to the customer's, which consisted of large-scale Domino workloads characterized by large mail files, documents with multiple attachments, frequent searching of documents, and full-text indexing (FTI). So IBM launched a 12-week benchmarking study to analyze workloads in these environments.

We discovered the following:

- ▶ Document count has a greater impact on performance than mail file size.
- ▶ Workloads using mixed file sizes behave like similar workloads using files of all one size.
- ▶ Upgrading from Domino R5 to 6 can improve performance by as much as 30%.
- ▶ Reducing the number of documents kept in the inbox reduces overall CPU usage, improves response time, and can dramatically improve startup/recovery performance.
- ▶ Full-text indexing costs little in terms of performance, but searching is expensive.

This paper details our process and key findings. The findings should be helpful to anyone responsible for sizing and implementing large-scale Domino servers. This paper also makes recommendations based on the findings and identifies additional resources to help readers size large-scale Domino environments on iSeries.

Benchmark methodology

A challenge in designing the large-scale Domino benchmark was the construction of a realistic test environment. Given the sensitivity of customer data, access to real data on large-scale workloads, database contents, and usage patterns was restricted. Therefore, the benchmark team derived workloads based on observed usage patterns of Domino customers, and simulated mail files based on limited access to a variety of Domino environments.

The results produced by this method might not hold precisely in every large-scale Domino environment. The team is confident, however, that their findings are useful as guidelines in sizing such environments and provide a foundation for developing further insight into the performance of workloads that had not been measured before.

Workloads

The team chose a predefined “6 Mail Routing” workload that is based on the experience of many Domino customers. Variations were run to measure ramp-up time and sustained activity at user levels ranging from 500 to 2000. Workload variables included mail file size, document size, FTI, document search, attachments in mail files, transaction logging, rate of mail delivery, and address book size.

Based on available customer data, the benchmark team defined the following user profile for the tests:

- ▶ Mail file size: 700M
- ▶ Document size: 25K
- ▶ Full-text indexing: On
- ▶ Transaction logging: On
- ▶ Document search: Included in simulation
- ▶ Response time required: Sub-second
- ▶ Transaction rate: Heavy
- ▶ Cluster: On

Tools

Workload was introduced to the test environment with the Domino tool *Server.load*, and 16 PC servers were installed to run this tool.

Server.load simulates the activity of Notes® clients on a Domino server by using PC clients that each represent up to 500 users. This is the same methodology used by IBM for other, smaller, Domino workloads. For each workload introduced, *Server.load* and iSeries data collectors provided statistics on CPU, disk, memory, system resource and network usage, and client response times.

Runs

Workloads were run to simulate operation at three different sites, each with different workload characteristics. Each run contained a ramp-up period during which sessions were attached to the system at the rate of one new session every two seconds. The ramp-up also simulated workload peaks typically resulting from failure recovery. After the ramp-up, each workload sustained a steady, defined level of activity for a specified time, then completed with a ramp-down that represented the removal of 500 users. Workload variables such as mail file size, indexing, or document size were changed from run to run. In order to isolate and

measure the impact of each change, usually only one variable was changed at a time. On average, it took 18 hours to set up and run each test.

Test environment

The team built an i830 server to the exact configuration of a model in operation at a customer site. This system was used as a control to help the team assess whether the workload generated in the test environment had characteristics similar to the work running in an actual site. Then an 8-way and a 24-way Model i840 servers were installed as target systems to test projected workloads. The details of these configurations are as follows:

- ▶ iSeries, 830-2349 (8-way 540MHz)
 - 20,910 Mail & Calendar Users (MCU) rating
- ▶ iSeries, 840-2461, 8 processors enabled
 - 40 GB memory
 - 5,051 GB of disk space (538 drives)
 - 25,930 Mail & Calendar Users (MCU) rating
- ▶ iSeries, 840-2461 24 processors enabled
 - 124 GB memory
 - 5,051 GB of disk space (538 drives)
 - 77,800 Mail & Calendar Users (MCU) rating

After the hardware environment was installed and operational, Domino 5.0.10 was installed and configured. Individual Server.load users were built and added to the Name and Address Book (NAB). Mail files of the required sizes were built to contain documents of specified sizes, composed of random English words. After a mail file was built to specification, it was copied for each of 2000 simulated users.

IBM benchmark process

Benchmarking provides users with accurate predictions of performance in targeted environments. Successful benchmark projects start with a Planning Session. At IBM benchmark centers, for example, representatives of the client company (including applications, operations, and user representatives), and often their Business Partner, meet with the benchmark center staff to discuss what questions are to be answered by the benchmark. The center staff uses an ISO9001 process to lead these discussions. After goals are defined, the client and benchmark center design a benchmark to achieve them. The benchmark design encompasses all phases of the application under study: composition and size of database; number of users; types of tasks performed by the users; frequency of task; target system configuration; and the tools used to simulate user workload. In typical engagements, a two-day planning session is sufficient to define goals and define a benchmark. The project team then spends several weeks to prepare for the study.

This time is used to capture or create databases of the appropriate size and structure, to work with the user community to understand the transactions and workload they introduce to the target environment, and to create the scripts used by workload simulation tools to drive work into the benchmark system. Data and scripts are then installed in the test environment and testing begins.

The first phase of testing is known as *calibration*. Its purpose is to build a baseline from which all other tests are measured. The calibration is often made from data collected at the home site from a production run of the workload being modeled. Calibration gives the team a chance to tune the benchmark design if any adjustments are necessary. After calibration is complete, the team runs the scenarios and tests defined in the benchmarking plan. The team collects and analyzes performance data at the end of each test, then runs the next scenario, often changing only one variable (such as number of users) so that the effect of each change

can be measured accurately. This process is repeated until the team has performance data for each variable and value specified in the benchmark plan.

Key findings

This section describes the areas to look at when tuning Domino.

Document count

The number of documents in a typical mail file has a greater impact on performance than the size of the file. For example, as shown in Table 1, tests run with 700 MB mail files containing 28,672 documents required two times the CPU required by 700 MB files containing fewer documents. This effect was seen only with large mail files, not in tests using 100 MB mail files.

Table 1 Correlation of mail file size and document count to CPU utilization

Mail file size	Document count	CPU utilization
700 MB	28,672	16%
700 MB	7,196	8%
100 MB	1,024	6%
100 MB	4,096	6%

Document count has its greatest impact on peaks such as startup and failure recovery. The impact on steady-state performance is lower. Figure 1 on page 5 illustrates the effect of document count on CPU usage.

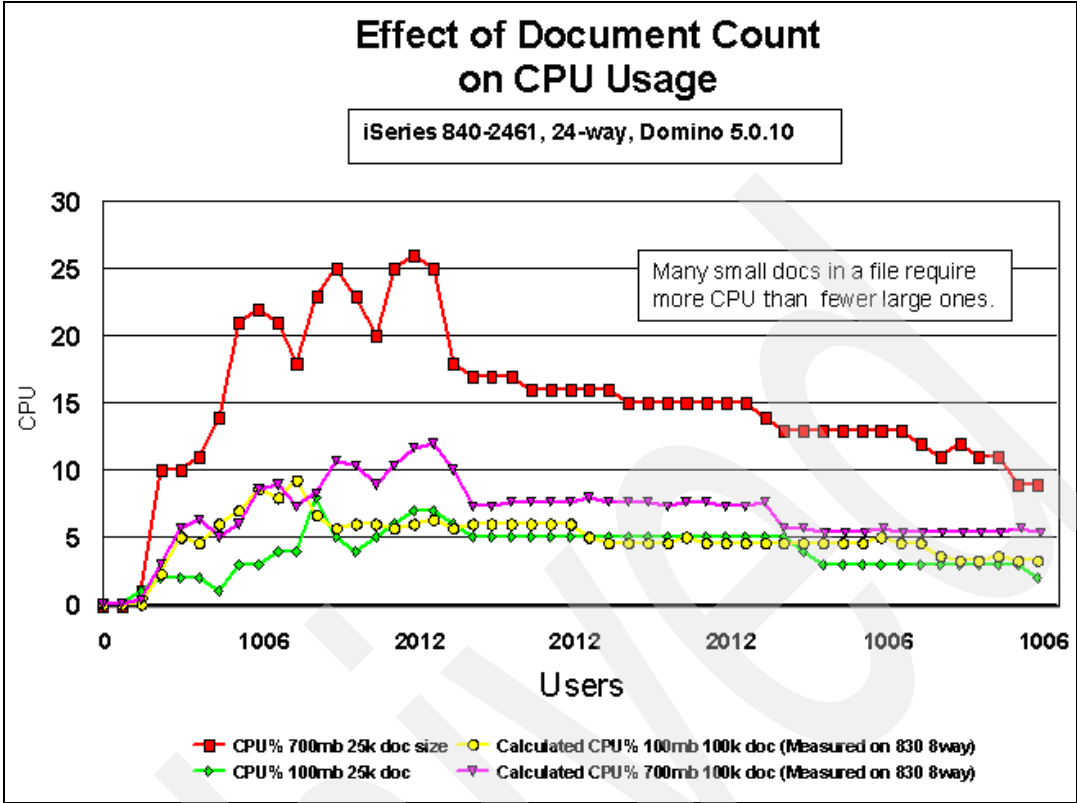


Figure 1 Effect of document count on CPU usage

Mail file size

The number of documents a mail file contains has a greater impact on performance than the size of the file. A file with many small documents requires more CPU usage than a file of the same size containing fewer large documents. Aside from peak states during work startup and failure recovery, file size has comparatively little impact until size approaches 700MB or more, as shown in Figure 2.

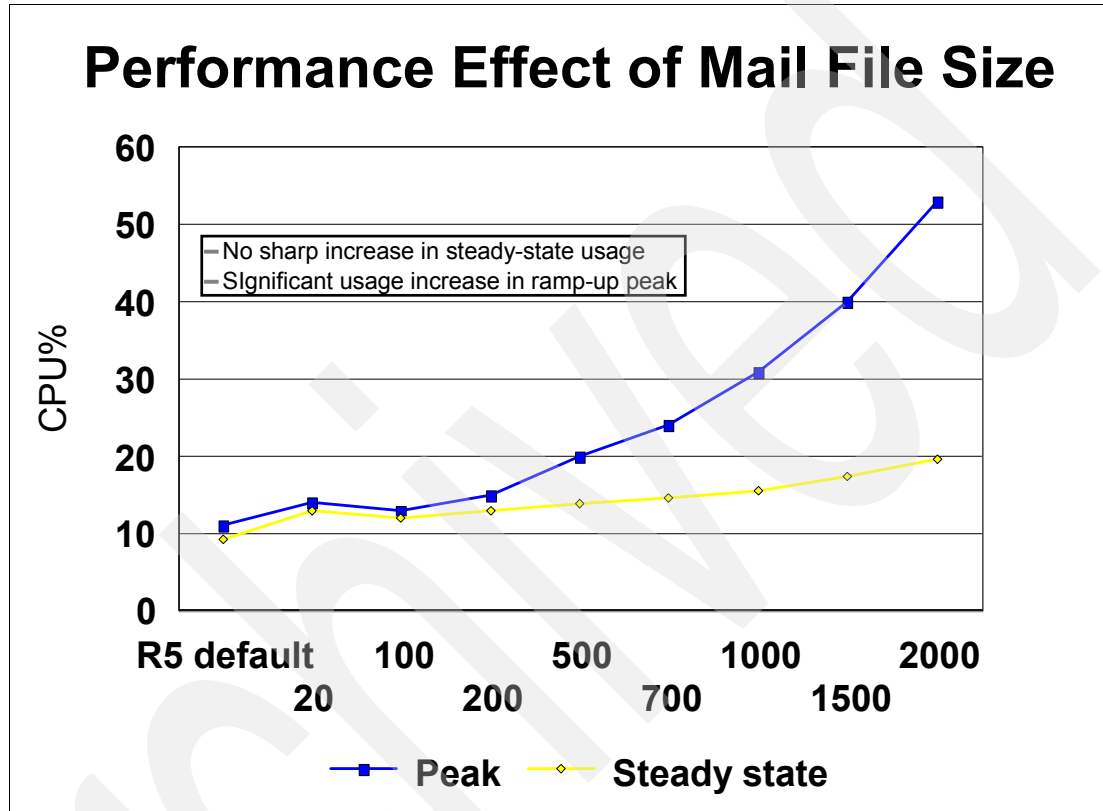


Figure 2 Effect of mail file size on CPU usage

Note: R5 default is the default mail file size for an R5 workload. File sizes are expressed in megabytes.

Mixed mail file sizes

Apart from ramp-up, mixing file sizes has little effect on performance. The steady-state, non-peak performance patterns of workloads using mixed files sizes resemble performance of similar workloads using files of all one size. See Figure 3.

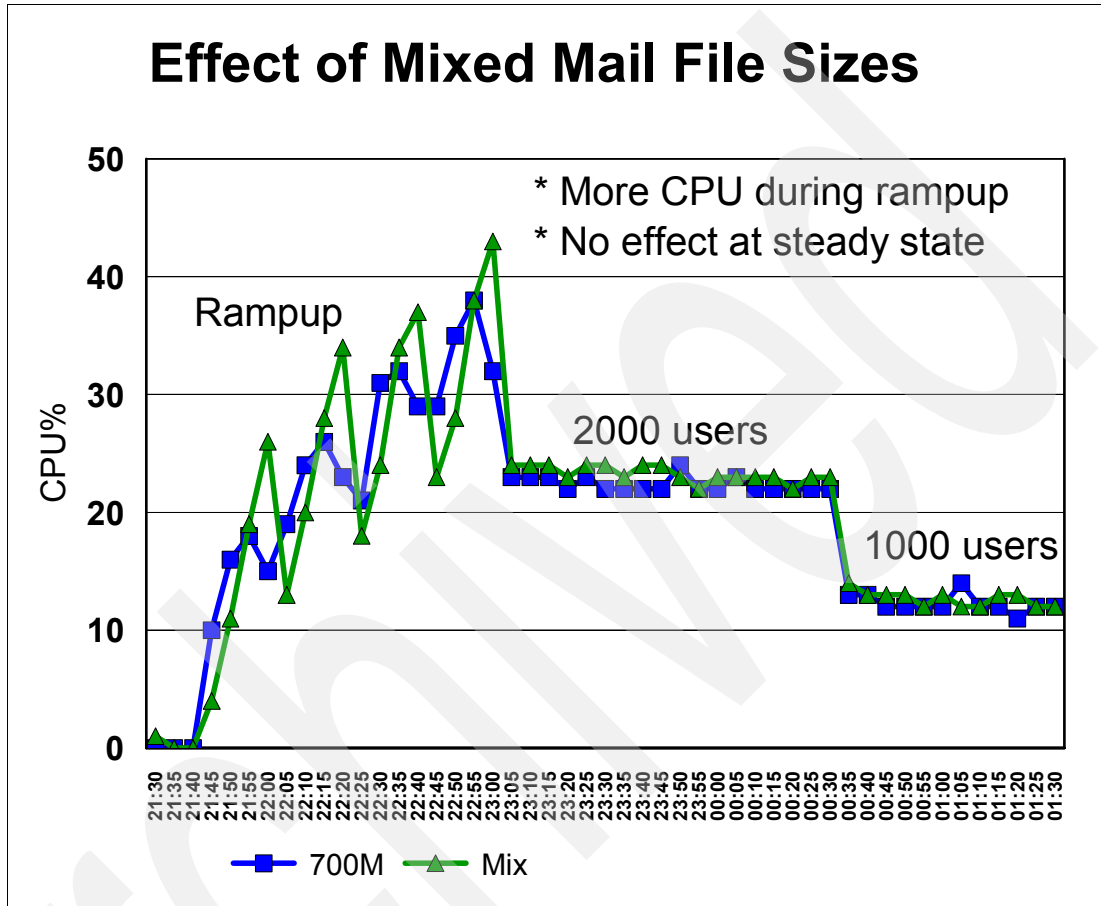


Figure 3 Effect of mixed file sizes on CPU usage

Impact of Domino upgrade

Upgrading Domino R5 to 6 improved steady-state CPU usage in our test environment by as much as 30%. The impact of the upgrade was lower at peak workloads, such as during session startup and failure recovery. Figure 4 illustrates the effect of the upgrade. Upgrading from 6.02 to 6.5 reduced CPU usage by another 6%.

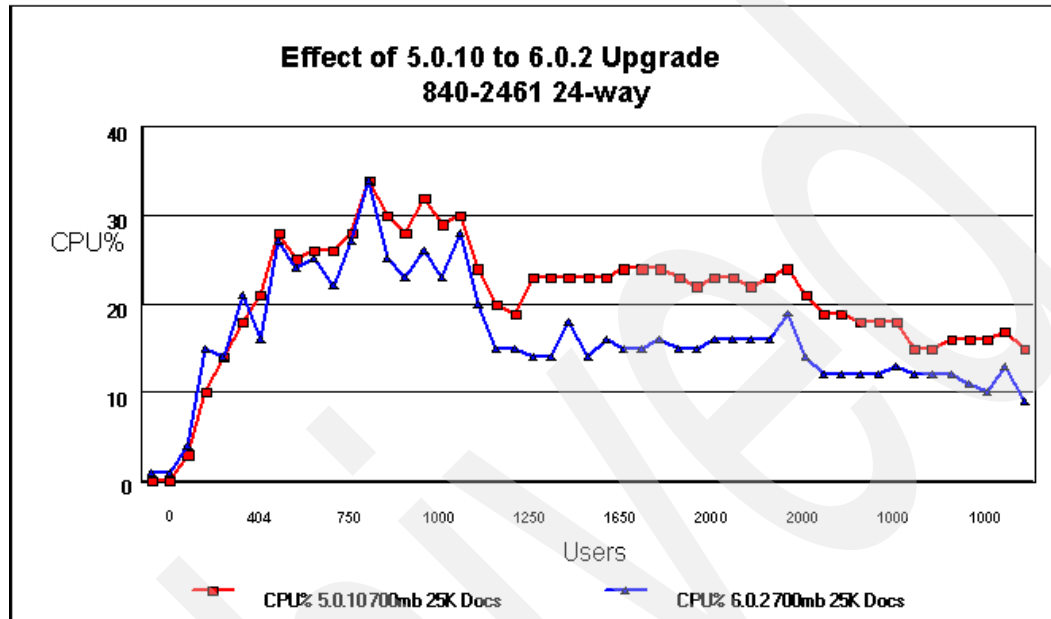


Figure 4 Comparison of performance running 5.0.10 and 6.0.2, including FTI, search, and attachments

Effect of storing documents in the Inbox

The Inbox is the most expensive container in a mail file. The Inbox folder, which contains all new messages received to a mail file, must be updated each time a user opens the file or clicks Refresh to see new mail. The more documents kept in the Inbox folder, the more expensive it is to refresh the view of it. Reducing the number of documents in the folder reduces the CPU and main storage required to update the view of it.

The effect on startup/recovery performance can be dramatic. As illustrated in Figure 5, storing 100% of the test documents in the Inbox created a usage spike during ramp-up. Reducing the number of documents in the inbox to 25% (all other documents being in a folder) virtually eliminated the spike. The performance benefit of eliminating the ramp-up spike occurs whether a Domino session is initiated normally or during failover recovery.

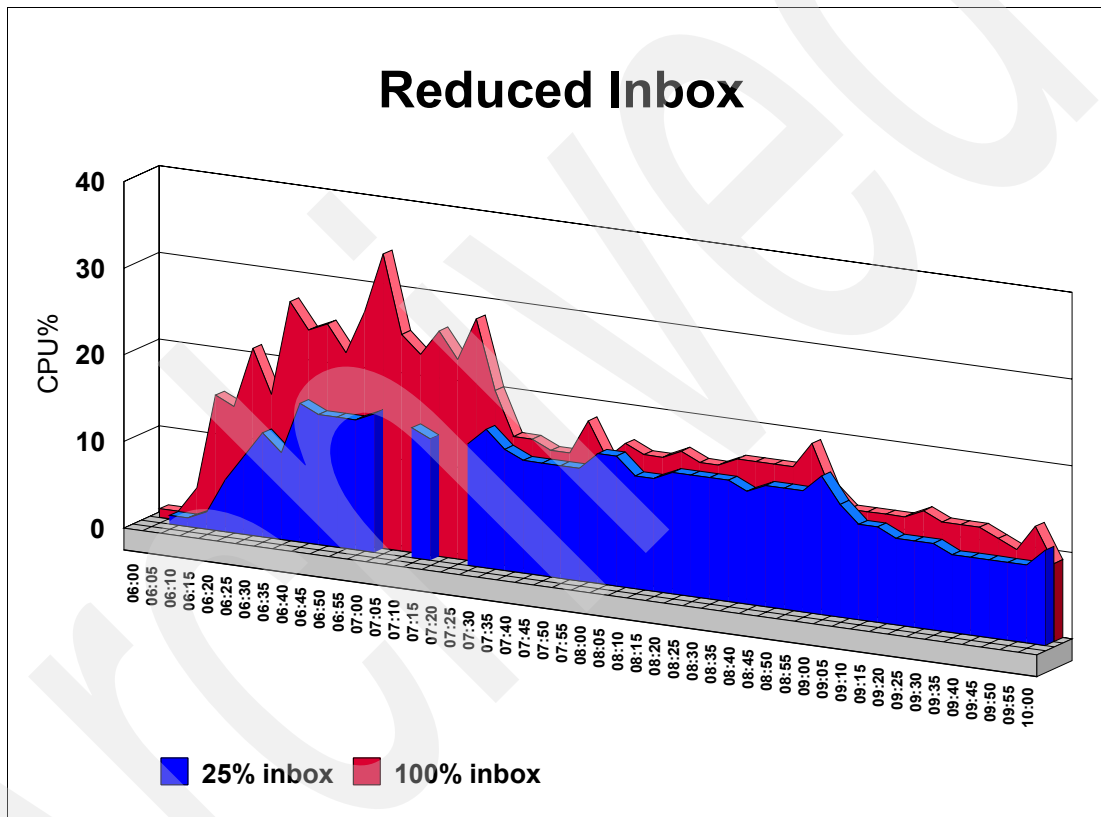


Figure 5 Impact of reduced number of documents in Inbox

Note: Spaces in the 25% graph indicate that the test was not run for those time intervals.

Full-text index and search

Full text indexes cost little in terms of performance, but searches are expensive. For example, adding full-text indexes raised CPU usage in our test environment from 13.1% to only 13.2%. Workloads requiring each user to search the Inbox once every 15 minutes, however, increased CPU usage by 20%. See Figure 6.

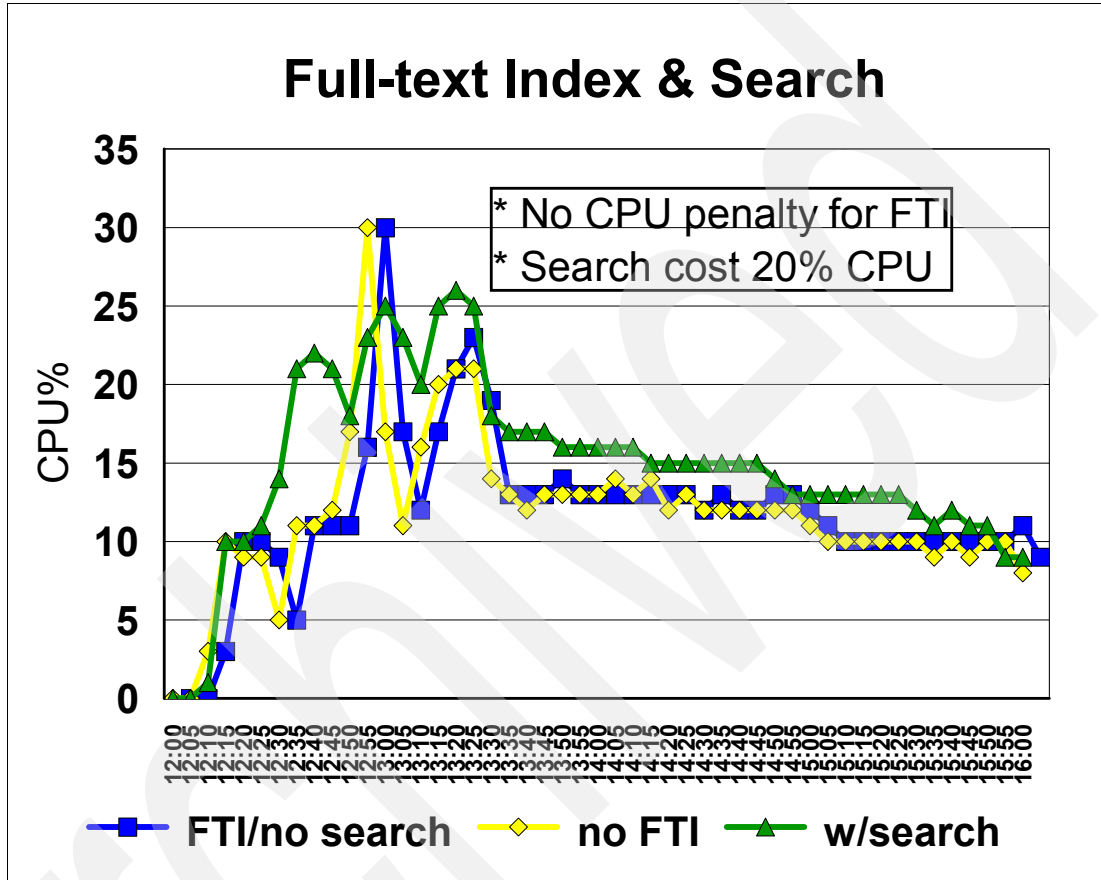


Figure 6 Full-text index and search

Impact of partition count

Each new partition requires additional system resources. For example, processing cycles and memory are required to support the 200 or so threads generated for each new partition. Our estimate for the test environment is that each new partition adds about 0.1% CPU and 192 MB of main storage overhead. More difficult to determine is how the workflow changes as the partition count increases. We assume that average hops per mail message increase as the number of partitions increases, because the likelihood that messages are now on different servers also increases.

Our study included one scenario with 12 partitions for all mail workload and another scenario with 87. We estimated that implementing 12 partitions instead of 87 would reduce main storage requirements by as much as 320 MB per system and reduce CPU usage by 2.6% per system. Additional CPU savings from reduced mail hop could approach 30%, depending on locality of mail senders and receivers.

Additional findings

In addition to the finding described above, our performance study revealed that:

- ▶ Transaction logging raised CPU steady-state usage 15%, from 15.9% to 18.4%.
- ▶ Large file sizes required more memory for fault recovery and paging.

Recommendations

The following recommendations are based on our key findings in the large-scale Domino workloads study. See “Appendix A: Solutions for managing documents” on page 13 for ways that IBM can help with your Domino performance tuning.

Keep current

Upgrading from Domino 5.0.10 to 6.0.2 improved performance in our test environment by as much as 30%. Further, upgrading from 6.0.2 to 6.5 improved performance another 6%.

Manage document counts

One of the quickest ways to improve the performance of a large-scale Domino environment is to reduce overall document counts and the number of documents stored in Inboxes:

- ▶ CPU usage for our 700 MB files containing about 29,000 documents was twice the usage of 700 MB files containing about 7,000 documents.
- ▶ Reducing the number of documents stored in inboxes from 100% to 25% reduced CPU usage at ramp-up by as much as 50%, and in steady-state by as much as 12%. Making a habit of filing documents in folders instead of the Inbox can dramatically improve peak performance and improve steady-state performance as well.

Appendix A lists tools available to help you manage document counts.

Manage partition counts

Reducing the number of partitions improves performance. Eliminate or consolidate partitions to reduce CPU usage required for routing, replication, administrative processing, and thread management.

Draw on available expertise to help size large-scale Domino environments

Sizing large-scale Domino environments is complex. The performance impact of various workload characteristics and their interrelationships are amplified in such environments. Based on our study, we recommend that customers seek assistance from IBM/Lotus® to help plan and size Domino iSeries environments that have one or more of the following characteristics:

- ▶ 4,000 or more mail users
- ▶ Mail file sizes of 500 MB or more
- ▶ Planned to reside on multiple iSeries servers
- ▶ Will replace 10 or more Intel® servers

Resources:

- ▶ IBM Redbook *Domino for iSeries Sizing and Performance Tuning*, SG24-5162
- ▶ Your local IBM representative, IBM Business Partner, or IBM technical support center

- ▶ Lotus Field Support Services:
<http://www.lotus.com/services/fss.nsf/>
- ▶ iSeries and AS/400® Benchmark Center:
<http://www.ibm.com/eserver/series/benchmark>

Archived

Appendix A: Solutions for managing documents

As the Domino benchmarking study showed, performance is affected dramatically by the number of documents users keep in their mail files and the number they store in Inboxes. Growth of mail files naturally occurs with increased activity, but growth is compounded in the absence of controls.

Some customers introduce voluntary controls, such as requesting that users remove mail after a certain time period, or keep file size to a specified maximum (300 MB, for example). Voluntary programs can control growth of mail files, but they entail the risk of loss of intellectual capital. To maintain file size within a specified limit, for example, users might inadvertently delete critical documents that could have been safely stored in a controlled, shared repository.

IBM offers a number of tools to help customers manage the growth of mail while reducing the risk of loss. Tools can help identify messages that have met retention periods, help manage archiving of messages for long-term storage, and help protect critical intellectual capital. Below are just a few of IBM's packaged solutions for managing Domino documents. Contact IBM or your IBM Business Partner for more information on these solutions.

IBM Lotus Domino 6

With IBM Lotus Notes® 6 and Domino 6 (ND6), there are new capabilities for managing mail in the enterprise. ND6 includes the Journaling feature that enables an organization to capture all incoming and outgoing messages into the system centrally. Additionally, ND6 offers the ability for administrators to enforce server-based or local mail archives through the new Policy facility. Spam filter options enable reduction of mail before it is added to mail files.

The ISSL Domino Mail Manager asset

IBM Software Services for Lotus (ISSL) offers an asset that provides an enterprise with a tool that can assist in managing the message store. Domino Mail Manager was designed to help enterprises manage their Domino mail servers by providing the capabilities to capture messages for archiving, purging messages that have met retention requirements, purging duplicate attachments from a single user's mail file, and purging large attachments from the mail file. This asset is sold as a Service offering to clients.

IBM Common Store for Lotus Domino

Common Store for Lotus Domino is part of the IBM Content Manager portfolio of products. Common Store provides the capabilities to enable users to store message content in the Content Manager solution from IBM. This includes single documents, file attachments, or entire folders. The archived content is also easily accessible from the IBM Lotus Notes client by users. Content can also be indexed and searched through the addition of the IBM Lotus Discovery Server. Additionally, the ISSL Domino Mail Manager asset also integrates with Common Store, allowing for automated message archiving.

IBM Tivoli Storage Manager

Tivoli® Storage Manager (TSM) is the IBM enterprise backup solution. TSM integrated with Tivoli Data Protection for IBM Lotus Domino enables organizations to capture the entire message store and provides storage to back up media such as magnetic tape and optical storage. TSM can also be integrated with IBM Common Store to allow for further offloading of Content Management data.

IBM Lotus Domino.Doc

The IBM Lotus Domino.Doc® software is an Enterprise Document Management system based on the IBM Lotus Domino platform. Using Domino.Doc in conjunction with an Enterprise Mail Management strategy enables an organization to deploy a highly robust content store for archived messages and key intellectual capital. Messages can be moved from the user's mail file and stored in the Domino.Doc store, allowing it to be indexed as well as providing for full text search capabilities. Domino.Doc can be augmented with Domino Mail Manager to provide for an automated solution for capturing messages and storing them in a searchable archive. Finally, TSM can be deployed to further off-load message content to magnetic tape and WORM media.

IBM My Attachments tool

The My Attachments tool provides a mechanism to identify and manage attachments in a user's mail database or any other database for which the user has edit capabilities. Attachments in selected documents can then be moved to a repository database of the user's choice, typically residing on the user's local hard drive. Compacting the local database frees space occupied by the attachments. Server-based mail files are compacted weekly and should not be initiated by end users. Because this application only stores a single copy of a particular attachment and can remove all occurrences of the attachment from the source database, considerable savings in disk space can be realized. Subsequent moves of the same attachment will delete the attachment from the analyzed database and point to the attachment copy already in the repository.

The team that wrote this Redpaper

This Redpaper was produced by a team of specialists working at the International Technical Support Organization, Rochester Center.

Lynn F Behnke is a retired Senior Program Manager. His IBM experience includes technical writing, publications management, and competitive analysis. He recently returned to IBM to lead development of a class on Competitive Selling. Lynn can be reached at <mailto:lfbroch@att.net>.

Beverly Cusato is an IBM/Lotus Service Manager with the World Wide Customer Advocacy Organization. Since she began working for IBM more than three years ago, Beverly has acted as an advocate and liaison for several Enterprise Clients within the support, development, and product groups across all products. As a Service Manager she works to ensure that IBM clients have world-class service and provide them with the information they need to succeed with the IBM/Lotus product family. Beverly is also a contributor to the Web-based training for WebSphere® Application Server 4.0.

Dan Frissora is an Advisory Software Engineer and a Project Manager at the iSeries Benchmark Center in Rochester, Minnesota. In his 25 years with IBM he has been involved with IBM Internal Applications Development, Expert Systems, AS/400 Performance Tools Development, AS/400 Systems, and Performance Management, and most recently the iSeries Benchmark Center. Dan recently returned from an assignment in Montpellier, France, where he helped initiate the EMEA iSeries Benchmark center in the IBM PSSC Montpellier.

Dave Johnson is an advisory software engineer at IBM in Rochester, Minnesota. He works on iSeries System Performance team focusing on Domino and Lotus collaboration products. His 19-year career with IBM has been spent mostly in performance-related work. He also worked in OS/400 development writing code for the Integrated File System (IFS). Dave can be reached at <mailto:davejohn@us.ibm.com>.

Joe Peterson is a Senior Software Engineer on the iSeries development team in Rochester, Minnesota. His focus is on performance of Lotus products.

Walter Scanlan is an Advisory Software Engineer working on Domino for iSeries. During his six years with IBM, he has been the worldwide technical leader for post-sales support of Domino on iSeries, responsible for education development and delivery, tools development, and providing best-in-class client service. During this time he was also an Adjunct Professor at Rochester Technical Community College. Prior to joining IBM, Walter worked for Computer Generated Solutions, an IBM Business Partner, working on AS/400 solutions for customers in the New York area.

Richard Smitherman is an Advisory Software Engineer and a Performance Specialist at the iSeries Benchmark Center in Rochester, Minnesota. He provides performance analysis and consulting and technical support for iSeries customer benchmarks. Before joining the iSeries Benchmark Center in 1997, he worked for IBM as an IT technical support specialist and team leader, a product planner for AS/400 office, communications, and system hardware, and a Systems Engineer in the New Orleans branch office supporting customers with System/3, System/32, System/34, and System/38™ systems.

Thanks to the following people for their contributions to this project:

Sue Ringwall
International Technical Support Organization, Rochester Center

Mike Ebbers
International Technical Support Organization, Poughkeepsie Center

Archived

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.

Trademarks

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

AS/400®

Domino.Doc®

Domino®

ibm.com®


IBM®

iSeries™

Lotus Notes®

Lotus®

Notes®

Redbooks(logo) ™

Redbooks™

System/38™

Tivoli®

WebSphere®

The following terms are trademarks of other companies:

Intel is a trademark of Intel Corporation in the United States, other countries, or both.

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