

# **Capacity Planning for IBM Netfinity Servers Using Lotus Server.Planner**

**by Clement Moy  
The Personal Systems Group  
IBM Corporation**

© Copyright 1998 Lotus Development Corporation. All rights reserved. Printed in the United States.

Lotus and Lotus Notes are registered trademarks and Domino and NotesBench are trademarks of Lotus Development Corporation. IBM, OS/2, AS/400, S/390, AIX and Netfinity are registered trademarks of International Business Machines Corporation. All other company and product names are property of their respective owners.

---

## Contents

---

Executive Summary .....	1
Introduction to Server.Planner .....	2
How hardware vendors generate Server.Planner data .....	2
Our Methodology for Generating Data .....	3
Netfinity configurations used with Server.Planner .....	3
NotesBench script selection .....	4
Generating valid NotesBench data .....	4
Interpreting performance data .....	5
How performance relates to system capacity .....	7
Using Server.Planner .....	11
Downloading Server.Planner and the Netfinity data .....	11
Examining the Netfinity Data Set .....	11
Building Queries .....	14
How Server.Planner interprets your query settings .....	14
How Server.Planner generates query results from vendor data .....	15
Example query settings and results .....	16
Categorizing users and workloads .....	17
Example Capacity Planning Exercise .....	19
A different interpretation of the data .....	20
A different way to represent user activity .....	20
Future directions .....	22

---

## Executive Summary

---

This white paper explains the methodology used by the IBM Personal Systems Group to generate data on IBM Netfinity servers for use with the Lotus Server.Planner capacity planning tool. It also suggests ways to use that data most effectively with Server.Planner.

Used together, Server.Planner and the accompanying data on Netfinity servers provide Lotus Domino customers with a sound, valid methodology for estimating actual server hardware requirements in their unique environments.

This paper covers the following topics:

- It includes instructions on how to download the Server.Planner software and the Netfinity performance data that IBM provides.
- Next, it explains how to read and interpret the Netfinity performance data.
- It also illustrates the process for making queries against the performance data. Both simple and complex examples are provided, to more clearly show the effects of various query options on the results that Server.Planner returns. The purpose of this section is to give Server.Planner users the benefit of our experience, so that their queries are as meaningful as possible.
- Finally, we provide an example set of capacity planning queries made by a fictitious service company, which plans to deploy a distributed network of Domino servers to its branch offices. This example discusses assumptions and decisions made by the fictitious customer, leading to the selection of the appropriate Netfinity server configuration for their capacity needs.

For a quick overview on how to use the tool, proceed to the section [“Using Server.Planner.”](#) To understand how the data sets are generated and why they can be used in planning your server requirements, continue with the next section.

---

## Introduction to Server.Planner

---

The Server.Planner capacity planning tool ships with Domino Server release 4.6.1 or later. It can also be downloaded free of charge from the Lotus Notes.Net Web site ([www.notes.net](http://www.notes.net)). Server.Planner is designed to assist Domino customers in purchasing server hardware, by allowing them to “run” hypothetical Domino/Notes workloads against a range of server configurations. Server.Planner compares the workload specified by the query with independently-certified NotesBench data supplied by the hardware vendor.

An important part of this equation is the data supplied by the hardware vendor, who must select appropriate workloads, system configurations and operating conditions in order to collect valid performance data. The validity of this data essentially determines the validity of the results Server.Planner generates in response to customer queries. The vendor’s choice of system configurations, operating environment and data sampling techniques are all important.

We in the IBM Personal Systems Group have taken great care to continually refine and improve our data gathering methodology, ensuring that customer queries concerning our Netfinity servers will return usable results. Additional explanations concerning our methods, as well as our most recently collected data, are available on the NotesBench Consortium Web page ([www.notesbench.org](http://www.notesbench.org)).

### How hardware vendors generate Server.Planner data

To enable customers to use Server.Planner in capacity planning for Domino Servers, we at the IBM Personal Systems Group, along with other hardware vendors, must first capture performance data on our servers, using special NotesBench tools and scripts. These results are collected and saved in a database, created using the Server.Planner “vendor data” template.

After a vendor has collected the performance data generated by running a set of NotesBench scripts on a specific server configuration, the vendor submits the data to the NotesBench Consortium’s auditing agency, KMDS Technical Associates, Inc., for review and certification. Successfully certified data receives an electronic stamp, which enables customers to make queries against it with Server.Planner using the “analyst” template.

---

## Our Methodology for Generating Data

---

This section describes the methods used by the IBM Personal Systems Group to develop Netfinity server configurations and generate meaningful, consistent NotesBench data against those configurations. It explains:

- The Netfinity server configurations the NotesBench scripts were run on, and why
- Which NotesBench scripts were used, and why
- How a valid range of data points was generated for use by Server.Planner

### Netfinity configurations used with Server.Planner

The IBM Personal Systems Group provides a range of Domino Server platforms to handle deployments from small offices to large data centers. Each specific server configuration for which we provide Server.Planner data targets a specific segment of this range. Naturally, it is possible to add optional components, such as memory and disk drives, so that the capacities of servers originally designed for different market segments would overlap. However, to maximize the value of our data, we have intentionally minimized the overlap in capacity among the systems we tested.

We chose Netfinity system configurations for NotesBench testing based on these requirements:

- **Acceptable user performance even at the maximum recommended level of system utilization.** To produce data suitable for capacity planning, we run our servers at the level of resource usage *we recommend to customers when they initially deploy a new server* (i.e., a “safe” level that allows for a good deal of unanticipated workload and some early growth). For example, when running the NotesBench script Mail and Shared Database (MailDB), which is designed to simulate medium-intensity messaging usage, we generate data only against server configurations that provide acceptable performance even when “CPU Utilization” reaches about 70%. For more information, see the section “Generating valid NotesBench data” below.
- **Optimal utilization of system resources.** The number of processors, amount of memory and number of disk drives on each system are carefully balanced to produce acceptable user performance, while maximizing the utilization of these upgradeable components. In other words, these system configurations are those which will give customers the most value for a given capacity requirement, while allowing room for cost-effective expansion.

- **Consistent data collection.** We collected Server.Planner data across various workload benchmarks using the same server configurations. This greatly improves the value of our data for comparative purposes, and eliminates any possibility of skewing the data across different workloads to inflate performance statistics. Customers can specify server workload using a mix of NotesBench scripts, and the data returned will be valid. If at any time during our testing, any measure of performance (user-perceivable performance, Probe Response Time or NotesBench Response Time) reached a threshold we considered unacceptable, it became the criterion for maximum recommended system utilization for that configuration and workload, regardless of CPU utilization levels.
- **Support for interpolation and approximation.** To support approximations based on mixed workloads for which we haven't collected data, we collect data at lower levels of resource utilization. For example, in cases where user-perceivable response time degrades below acceptable levels before CPU Utilization drops below 70%, we collect Server.Planner data for CPU Utilization levels down to 50% or less. In other words, we always favor acceptable user response time over total system utilization.

## NotesBench script selection

The IBM Personal Systems Group has chosen five NotesBench “user scripts,” which provide performance information valid for users of Notes clients: Mail-Only, Mail and Shared Database, Groupware, Calendar and Scheduling, and Web-Walker.

The IBM Personal Systems Group will, of course, take into account market changes and customer feedback when we select NotesBench scripts for generating Server.Planner data.

## Generating valid NotesBench data

Benchmark testing scripts cannot accurately represent every customer's unique, real-world environment. To account for variations that an actual Domino server may encounter (such as additional server tasks), our tests attempt to simulate a user load that keeps CPU utilization on the server in the 50% to 70% range, while all users are executing the NotesBench “Mail and Shared Database” (MailDB) script.

MailDB puts a much lighter CPU load on a server than the script “Groupware.” On the other hand, MailDB puts a much heavier load on the disk subsystem than Groupware. Had we used Groupware to determine the server configuration, you would see more CPU or fewer disk drives in the configurations.

While our basic assumption is that much of the work done by Domino servers will be mail-related, our Netfinity test servers are configured to produce acceptable user performance across a range of usage profiles. Thus, you may find that the volume of disk storage in these server configurations exceeds your anticipated needs, particularly if you estimate your requirements solely on a minimum storage allowance per user, for example.

Likewise, with the addition of RAID subsystems, more drives improve disk subsystem performance. Keep this in mind when selecting the size of your disk drives. Disk drive manufacturers usually offer two to three storage sizes in a family of devices with identical performance. The intermediate sizes often offer the lowest cost per unit of disk storage. When appropriate, consider choosing more of the smaller drives instead of a few larger drives. You may achieve better RAID performance while benefiting from the lower cost of the smaller drives.

**Note:** We conduct our Server.Planner analysis under the strictly defined conditions required for a NotesBench certification. This is essential for you to compare our NotesBench reports with those of other vendors. However, it means that our test servers run only those services specified by NotesBench. If you run additional applications or services in your environment, these will add to the server workload and possibly reduce the number of concurrent users supported.

## Interpreting performance data

The performance results that Server.Planner generates reflect a range of metrics. Two metrics related to the response time experienced by users are **Average NotesBench Response Time** and **Average Probe Response Time**.

**Average NotesBench Response Time** is the average of response times of all transactions executed by each simulated user during the test, as measured at the test driver. It is a good overall estimate of server performance for a specific workload. Average NotesBench Response Time can be used to compare different servers for a specific workload.

However, each script simulates an additional user, called the **Probe**. The Probe continuously opens and closes a high-contention database shared by all the simulated users. **Average Probe Response Time** reflects the average performance seen by the Probe over all the file open-and-close operations it performs. Average Probe Response Time more accurately reflects what an individual user will perceive in terms of “end user” workstation response time. It can be used to compare server performance across the different workloads.

We define acceptable Probe response times to be under 5 seconds; we further divide this range into fast, medium and slow. Our data for servers running MailDB at 50% to 70% CPU Utilization rates generally produce an average Probe response time in the “fast” range.



Figure 1 below shows the relationship of Average Probe Response Time, Average NotesBench Response Time and CPU Utilization against number of MailDB users on the system. Figure 2 below shows the same Response times and CPU Utilization plotted against the metric NotesMark, a system throughput measurement defined as the number of Domino Server transactions completed per minute.

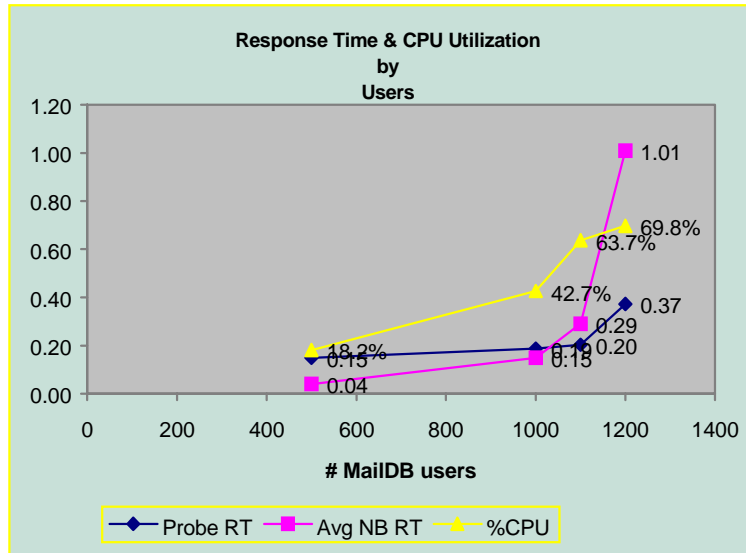


Figure 1

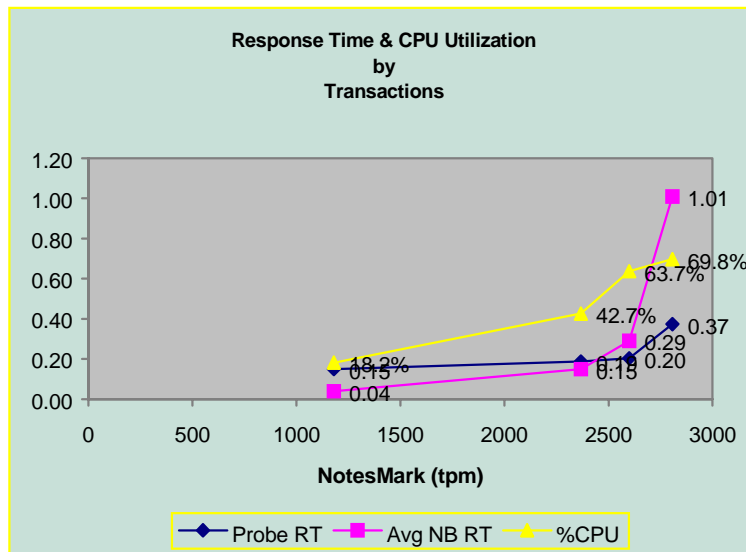


Figure 2

Note that Average CPU Utilization as shown in Figure 1 is a good indicator of the optimal number of users for this server configuration given this workload. The lines plotted on these two charts are essentially identical because the server under test is operating at a usage level where every additional user results in a proportional increase in useful work completed by the server.

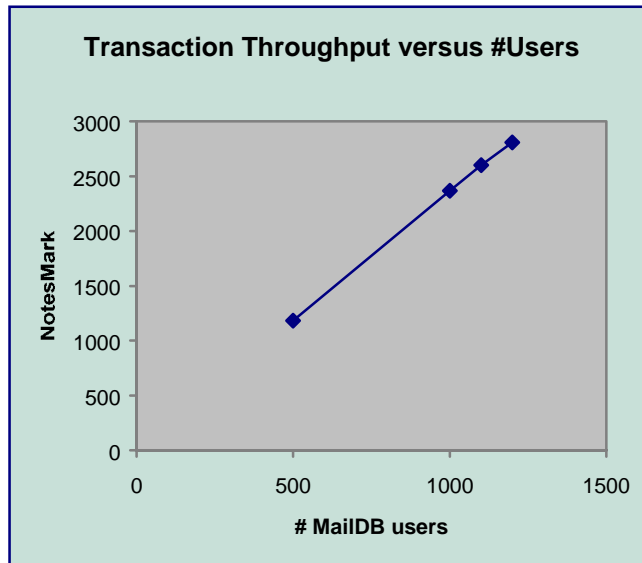


Figure 3

Figure 3 illustrates this linear relationship between NotesMarks and concurrent MailDB users. It verifies a proportional increase in server throughput with each incremental increase in the number of users, across the data points measured.

Referring again to Figures 1 and 2: Although the server under test was able to produce more work for each incremental increase in the number of connected users, the User Response time rises more sharply after the Average CPU Utilization goes above 50%. While each user still experiences fast response time, as both Probe Response Time and NotesBench Response Time indicate, up to about 70% CPU utilization, response time reaches the “knee-of-the-curve” at around 50% CPU Utilization.

Internally, the Server.Planner “analyst template” uses the Average Probe Response Time from these data points to determine the level of system capacity used for each type of workload. In addition to the value of the average Probe Response Time, performance analysts can review the distribution of response times to determine whether a system is approaching its knee-of-the-curve, as response times will be more variable as system resources are stressed.

## How performance relates to system capacity

System capacity is not a function of Average CPU Utilization alone, of course. The percentage of available memory used and the percent disk volume used are examples of other important system capacity measures. In addition, internal software “resources” such as the maximum number of semaphores supported, or the number of resource locks available, also affect actual system capacity. Some operating system or application parameters can function as the limiting factor in system capacity. You can review the details of the published NotesBench reports for tips on some parameters that we optimized for these specific platforms.

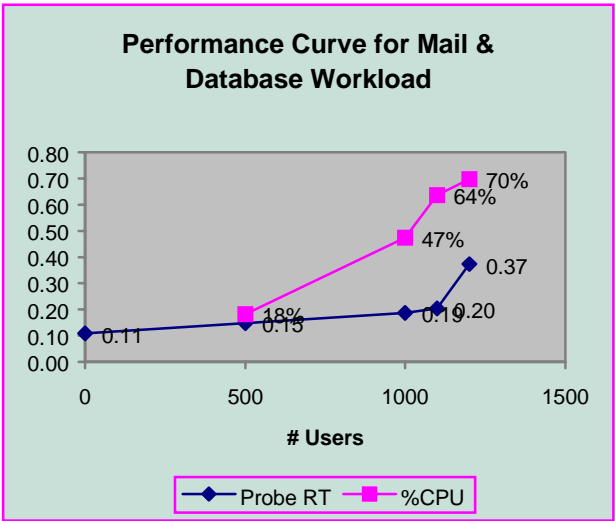


Figure 4: Performance Curve for Mail & Database Workload

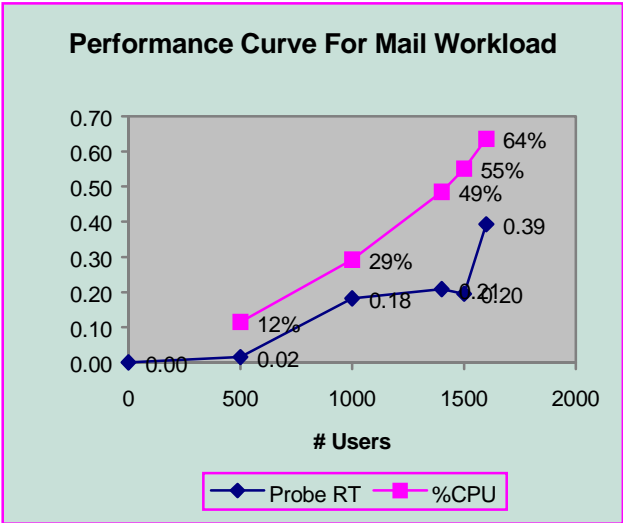


Figure 5: Performance Curve for Mail Workload

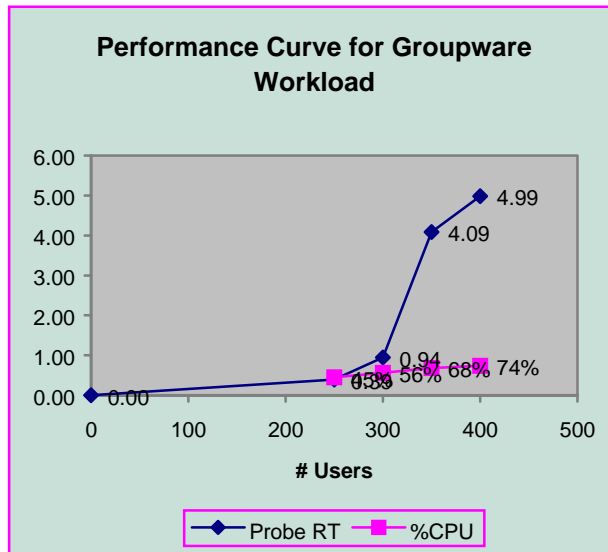


Figure 6: Performance Curve for Groupware Workload

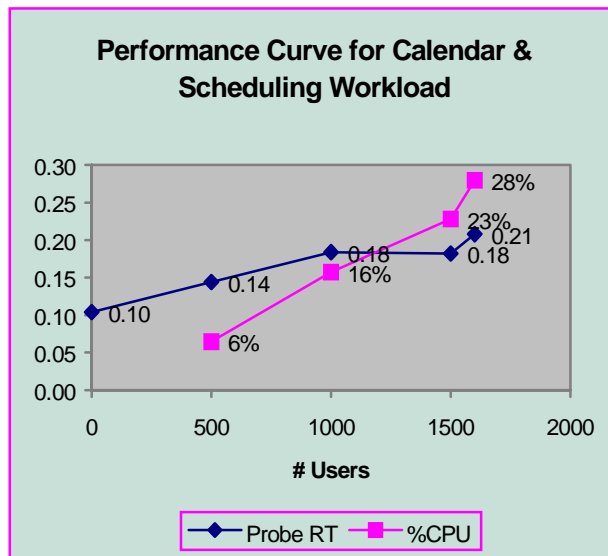


Figure 7: Performance Curve for Calendar & Scheduling Workload

Because of differences in system resource usage among these NotesBench scripts, CPU utilization is not always the limiting factor in determining the maximum number of users a configuration can support. For example, for Groupware users (Figure 6), the Probe Response reaches one second before CPU Utilization reaches 70%. For 300 Groupware users, Probe Response is 0.94 second, NotesBench Response is 0.26 second, and CPU Utilization is only 55%.

By contrast, consider our Server.Planner data (Figure 4) for 1,200 MailDB users, Probe Response is 0.37 second, NotesBench Response is 1 second and CPU Utilization is 70%. Thus, 300 groupware users generate approximately the same overall level of system utilization as 1,200 MailDB users, at least for this PC Server 704 configuration.

The same effect can be seen in other scripts. For example, this same server can support 1,600 concurrent mail-only users (Figure 5). In fact, for 1,600 Mail users the Average Probe Response Time is 0.2 second, Average NotesBench Response Time is 0.006 second and Average CPU Utilization is only 64%.

Our data for the Calendar & Scheduling (C&S) script (Figure 7) show that 1,600 C&S users result in an Average Probe Response Time of 0.2 second, Average NotesBench Response Time of 0.006 second and Average CPU Utilization of only 28% for the PC Server 704. We chose not to generate data for a larger number of C&S-only users because we assumed that a Notes user will generally use e-mail, but not necessarily use C&S; whereas a C&S user will also be an e-mail user.

As you can see from this brief exercise, the effective user capacity of a Domino server is dependent on the type of workload and the mix of usage in your specific environment. Thus, it is important to understand your user community in applying these results to your planning needs.

---

## Using Server.Planner

---

This section explains the basics of how to download Server.Planner and begin using it. For more information, see the comprehensive and invaluable *Server.Planner User Guide* included with the software.

### Downloading Server.Planner and the Netfinity data

To download Server.Planner and the data pertaining to IBM Netfinity servers, perform these steps:

1. Connect to the NotesBench Consortium Web site at [www.notesbench.org](http://www.notesbench.org).
2. Click the “NotesBench Reports & Server.Planner Data sets” button. (You’ll need to be a registered user.) This serves up the NotesBench Report and Server.Planner Data page. Click “Server.Planner Vendor Data Sets.” Find the document “Server.Planner Data – IBM Personal Systems Group.”
3. Follow the instructions to read about our methodology and download the executables for the data set, application templates and the *Server.Planner User Guide*. The files and the Domino database file are all contained in self-extracting .EXE files.
4. Run the extract executables. The database file `cpibmpc1.exe` contains the latest Server.Planner data for IBM PC Servers and Netfinity Servers, `cpibmpcc.nsf`. The `dspusgnsf.exe` creates the User’s Guide. The other executables create templates for “Vendor” (`dspv.ntf`), “Analyst” (`dspa.ntf`) and “Decision Maker” (`dspd.ntf`) applications. (Updates to these templates are maintained on <http://www.notes.net>. Check this site regularly for the latest version of Server.Planner.)
5. Copy these files to your `notes\data` subdirectory. If you want to use the charting capabilities of the templates, you should have a copy of Lotus Components 1.2 or later. Start or restart your Notes client and you are ready to use Server.Planner.

### Examining the Netfinity Data Set

Before you use Server.Planner to direct capacity planning queries against the Netfinity data set, it’s helpful to take a brief look at the raw data. From your Notes desktop, choose File - Database - Open to open the database “IBM PC Co. Server.Planner Data.”

The first time you open the database, you’ll see the Machine Identification view. Each document in this view describes a Netfinity system for which Server.Planner data is available; for example, a Netfinity 3500 System with a 233MHz Pentium II CPU and 320MB of memory, etc., running Domino 4.6 on Windows NT 4.0. YES appears in the Signature field to indicate that the NotesBench certification agency, KMDS Technical Associates, Inc., has certified the data.

Once you've perused the specifics about a system of interest, it's a good idea to take a look at the data collected on it. To do so, click the Vendor Data button on the left panel to select the Vendor Data view. For the IBM Personal System Server.Planner database, you'll see a view similar to that shown in Figure 8 below. The view consists of documents describing various data collection points; sorted first by system configuration tested, and then by the NotesBench script run.

Machine ID	Script Run	# Users/Spokes	Avg Probe	Fast	Medium	Slow	Actions/Min	Bytes/Min	Signature
NF3500-233MHZ1P6-320MB	Calendar & Scheduling	200	0.179	100%	0%	0%	0	0	Yes
	Calendar & Scheduling	400	0.175	100%	0%	0%	0	0	Yes
	Calendar & Scheduling	600	0.175	100%	0%	0%	0	0	Yes
	Calendar & Scheduling	700	0.21	100%	0%	0%	0	0	Yes
	Groupware #1	100	0.864	81%	18%	1%	14	0	Yes
	Groupware #1	200	2.301	5%	90%	5%	79	0	Yes
	Groupware #1	300	3.679	2%	49%	31%	114	0	Yes
	Mail	200	0.173	100%	0%	0%	0	0	Yes
	Mail	400	0.174	100%	0%	0%	0	0	Yes
	Mail	600	0.179	100%	0%	0%	0	0	Yes
	Mail	700	0.77	83%	12%	3%	0	0	Yes
	Mail & Database	200	0.193	100%	0%	0%	0	0	Yes
	Mail & Database	400	0.173	100%	0%	0%	0	0	Yes
	Mail & Database	600	0.212	99%	1%	0%	0	0	Yes
NF5500-350MHZ2PII-1024MB	Web Walker	100	0.154	100%	0%	0%	0	0	Yes
	Web Walker	200	0.178	100%	0%	0%	0	0	Yes
	Calendar & Scheduling	1	0.143	99%	1%	0%	0	0	Yes
	Calendar & Scheduling	500	0.171	100%	0%	0%	0	0	Yes
	Calendar & Scheduling	1000	0.177	100%	0%	0%	0	0	Yes
	Calendar & Scheduling	1500	0.205	99%	1%	0%	0	0	Yes
	Calendar & Scheduling	2000	0.191	99%	1%	0%	0	0	Yes
	Calendar & Scheduling	2400	0.198	99%	1%	0%	0	0	Yes
	Database	1	0.173	100%	0%	0%	0	0	Yes
	Database	100	0.176	100%	0%	0%	0	0	Yes
	Database	300	0.338	99%	0%	0%	0	0	Yes
	Database	500	0.576	94%	4%	0%	0	0	Yes
	Database	700	1.097	88%	4%	0%	0	0	Yes
	Database	800	3.643	54%	20%	1%	0	0	Yes
	Groupware #1	1	0.156	100%	0%	0%	0	0	Yes
	Groupware #1	50	0.32	100%	0%	0%	12	0	Yes
	Groupware #1	100	0.362	99%	1%	0%	24	0	Yes
	Groupware #1	200	0.417	97%	3%	0%	46	0	Yes
	Groupware #1	300	0.498	95%	4%	0%	67	0	Yes
	Groupware #1	400	0.644	83%	16%	0%	90	0	Yes
	Groupware #1	500	0.953	67%	29%	0%	112	0	Yes
	Groupware #1	600	0.943	63%	28%	0%	133	0	Yes
Groupware #1	700	2.204	43%	46%	2%	155	0	Yes	
Mail	1	0.154	100%	0%	0%	0	0	Yes	
Mail	100	0.142	99%	1%	0%	0	0	Yes	
Mail	500	0.173	100%	0%	0%	0	0	Yes	
Mail	1000	0.18	100%	0%	0%	0	0	Yes	

Figure 8

As you can see, for the Netfinity3500 server, the maximum numbers of users simulated for capacity planning purposes is: 600 MailDB users, 700 Mail users, 300 Groupware users, and 200 WebWalker users.

This initial view provides a wealth of information for the performance analyst. For example, for the Netfinity 3500, from 200 to 300 Groupware users, there is a change in the distribution of probe response times from mostly medium to medium/slow. This is an indication that this server may be reaching its peak performance for this workload. Compare this to other servers in the dataset, e.g., the Netfinity 5500 or Netfinity 7000, which show ample capacity for additional users on this workload.

You can view this information by selecting the workload and then clicking the “Graph Performance” and “Graph Distribution” buttons, when you have Lotus Components installed. Figure 9 shows an example.

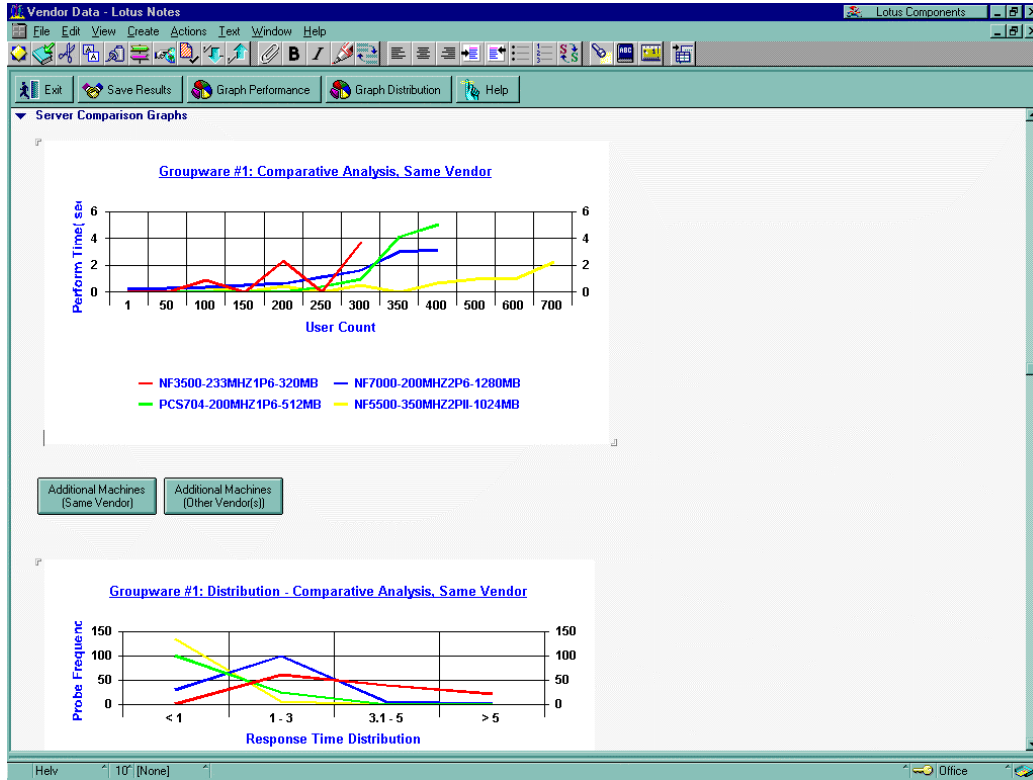


Figure 9

Experienced Domino administrators may wish to look at the performance characteristics of a few workload scripts to establish a range of users that fits with your experience. For example, “groupware” users may be similar to your “power user” community and may be compared with “mail & database” to provide a range for your capacity estimates. The scripts are described in more detail in the *Server.Planner User Guide*.

Whether you choose to use the maximum number of users simulated for each script or a lower user number depends on the level of system resources you want to keep in reserve as well as the level of response time degradation you are willing to accept during times of high unplanned usage.

The raw data provides an idea of a server’s capacity for a given number of users and a single type of usage. To estimate a server’s capability to handle a mix of user activity and numbers of users that reflects your actual environment, you need the powerful query capability of the Server.Planner “analyst” tool.



## Building Queries

To create a query, you need to create a file based on the “Server.Planner: Analyst” template. When you “Create Query on Data,” you’ll present queries to Server.Planner via the “User Task Information” form, shown in Figure 10 below.

In Figure 10, the analyst is interested in finding servers that can support 1,000 users. User activity is defined in User Profiles. The analyst has specified that the Mail-Only script will represent the work habits of 30% of the 1,000 users; the Mail and Shared Database script represents another 20%; and Calendar & Scheduling script represents 50% of the workload.

The results of this comparatively simple query indicate that either of the two Netfinity systems listed in the Results section at the bottom of Figure 10 meet the query’s criteria. Of course, their cost and extra capacity will differ.

The screenshot shows the 'Query on Data - Lotus Notes' application window. The 'User Task Information' section is expanded, showing three user profiles. Profile #1 is set to 1000 total users, while Profile #2 and Profile #3 are set to 0. The 'Server Analysis Results' section is also expanded, showing a table of results.

**User Task Information:**

User Profile #1:		User Profile #2:		User Profile #3:	
Mail (Low)	30 %	Mail (Low)	0 %	Mail (Low)	0 %
Mail & Database (Medium)	20 %	Mail & Database (Medium)	0 %	Mail & Database (Medium)	0 %
Groupware (High):	0 %	Groupware (High):	0 %	Groupware (High):	0 %
Database:	0 %	Database:	0 %	Database:	0 %
Calendar & Scheduling:	50 %	Calendar & Scheduling:	0 %	Calendar & Scheduling:	0 %
Web Walker:	0 %	Web Walker:	0 %	Web Walker:	0 %
Web Buyer:	0 %	Web Buyer:	0 %	Web Buyer:	0 %

Total Number of Active Users: 1000

Profile #1 Total Users:  Profile #2 Total Users:  Profile #3 Total Users:

Query Summary:

**Server Analysis Results**

Results:

Vendor:	Machine ID:	Description:	Response Time:	CPU/#:	Platform:
IBMPCCC	PCS704-200 MHZ1P6-512M B	PCS704 with 1 CPU (P6/200MHz) and 512MB RAM	Fast	Pentium Pro (P6)/200 x 1	NT 4.0
IBMPCCC	PCS325-233 MHZ1P6-384M B	PCS325 with 1 CPU (Pentium II/233MHz) and 384 MB RAM	Fast	Pentium II/233 x 1	NT 4.0

Enter optional descriptive text. Appears in View Pane.

Figure 10

## How Server.Planner interprets your query settings

One of the keys to using Server.Planner effectively is to get a feel for how the algorithm it uses to interpret your query settings (% Output Tolerance, Response Time and % Response Fall Within Range”) affects the query results. Referring back to Figure 8: Notice that for the Netfinity 3500 at the 300 user level, Server.Planner considers only 2% of the Groupware Probe samples to be “fast,” while 49% are considered “medium.” Based on experience, Server.Planner would not judge this server capable of supporting 300 Groupware-B users with a “fast” or medium probe response time. (For example, with default query settings of “medium” response time, 10% query tolerance and 75% of responses within range, Server.Planner actually judged our Netfinity 3500 configuration capable of supporting 222 Groupware-B users.)

Until you get an experiential sense for how the Server.Planner algorithm works, build your queries with more relaxed settings. For example: Specify slow response time, rather than medium or fast; this will accumulate all data points (slow, medium, and fast) in generating the result set—essentially nullifying the option “% Response fall within range” on the result. If a server can handle the specified level of system usage at a faster Response Time, it will be displayed as part of the query result. Likewise, specify a higher “% Output Tolerance,” like 50%. This changes a query specifying 1,000 Mail users, for example, to effectively mean at least 500 or more Mail users.

When you get too many systems satisfying a query, then start restricting the query options to make Server.Planner more selective. Consider a query specifying: 1,200 MailDB users; a Slow Response Time; 75% of responses within range; at 50% Query Tolerance. Six servers satisfying these criteria are the PC Servers 704, 330, and 325; and the Netfinity 3500, 5500, and 7000. Response Times for all servers are Fast. At Query Tolerance = 25%, all the servers except the Netfinity 3500 still satisfy the query. At a Query Tolerance of 10%, only the PC Servers 704 and 330 and the Netfinity 5500 and 7000 satisfy the query. At 0% tolerance, only the PC Server 704, the Netfinity 5500 and 7000 satisfy the query.

## How Server.Planner generates query results from vendor data

In general terms, Server.Planner uses the Average Probe Response at each data point to estimate how much of a server’s capacity is utilized by the number of users supported at that data point. In general, Server.Planner depends on measured data points to return query results. The more contiguous data points for every possible number of users are available, and/or the more data points spread out over regular intervals the vendor has supplied, the better Server.Planner’s estimate will be. In the future, Server.Planner will be enhanced for better interpolation of results.

For any specific NotesBench script, a higher number of supported users incur a higher level of server resource usage, eventually resulting in a higher response time. Take the example of the Groupware data for the PC Server 704, shown in Figure 6. Numbers of users and corresponding Average Probe Response Time for this script and server configuration are summarized in Figure 11.

<i>Number of supported Groupware users</i>	<i>Response Time (seconds)</i>
250	0.5
300	1
350	4.1
400	4.99

*Figure 11*

According to this data, the PC Server 704 as configured could barely support 400 Groupware users with acceptable Probe Response Time; it will not adequately support additional users beyond 400. Yet, if only 300 Groupware users were specified, one can see that conceptually there appears to be considerable additional capacity remaining. At 350 users, the server is much closer to capacity, but there still seems to be room for more users.

However, because of the data available, if a query specifies 351 Groupware users, Server.Planner rounds that number up to 400 users. Even if a query specifies 351 Groupware users and 10 Mail users, the 351 Groupware users will be treated as 400 Groupware users, leaving no reserve server capability to support the additional 10 Mail users (which in reality the server could easily support).

Likewise, Server.Planner treats 301 Groupware users like 350 Groupware users for this configuration, resulting in an over-estimate of the resources required to support that user profile. With a complicated query involving many user scripts, the impact of this “over-estimation by rounding up” can be quickly compounded. This may explain why analysts sometimes see “No Query Results Found” in response to a query that should be satisfied by one or more of the servers for which data has been provided. As mentioned in the previous section, the analyst can counter this effect by using a higher “% Output Tolerance” value.

## **Example query settings and results**

Figure 12 below illustrates a number of simple query examples, which we have applied to the Server.Planner dataset published by the IBM Personal Systems Group prior to September 28, 1998. Some of the examples illustrate the effect of different parameter settings on % Query Tolerance; Response Time and % Response fall within range.

The following servers are abbreviated in the table: PCServer 325 (**325**), PCServer 330 (**330**), PCServer 704 (**704**), NetFinity Server 3500 (**NF3500**), Netfinity Server 5500 (**NF5500**), and NetFinity 7000 (**NF7000**).

<i>Script/users</i>	<i>Script/users</i>	<i>Script/users</i>	<i>% Output Tolerance</i>	<i>% Response fall within Range</i>	<i>Response Time</i>	<i>Servers satisfying query</i>
Mail 700	Groupware 200	Cal & Sch 100	50	90	medium	704, NF5500, NF7000
Mail 700	Groupware 200	Cal & Sch 100	100	75	medium	704, 325, NF3500
Mail 700	Groupware 100		50	75	medium	704, 325, 330, NF3500, NF7000
Mail 700	Groupware 200		50	75	medium	704, NF7000
Mail 700	Groupware 100	Cal & Sch 100	50	75	medium	704, NF5500, NF7000
Mail 700		Cal & Sch 100	50	75	medium	704, 325, 330, NF3500
Mail 700	Groupware 100		25	75	medium	704
Mail 700	MailDB 200	Cal & Sch 100	50	50	medium	704, 325, NF3500
Mail 700	MailDB 200	Cal & Sch 100	100	50	slow	704, 325, NF3500
Mail 700	MailDB 200	Cal & Sch 100	25	50	slow	704, 325, 330, NF7000
Mail 700	MailDB 200	Cal & Sch 100	50	95	medium	704, 325, NF3500
Mail 100	Groupware 100		25	75	medium	704, 3500

Figure 12

## Categorizing users and workloads

Server.Planner allows you to represent user workloads in up to three different groups or user profiles. The work habits of each group can be described in terms of different percentages of time spent in activities represented by the various NotesBench user scripts.

Figure 13 shows an example query in which 1,000 users are classified into three different groups consisting of: 340, 350 and 310 users. Respectively, Profile 1, Profile 2 and Profile 3 represent the activities of the three groups. The 340-person group represented by Profile 1 spends 88% of its cumulative time on Mail activity and 12% on WebWalker-like activity. For this query, Output Tolerance is set to 50%, Response Time is set to medium and Response in Range to 75%. The results indicate that PC Server 704 matches the query.

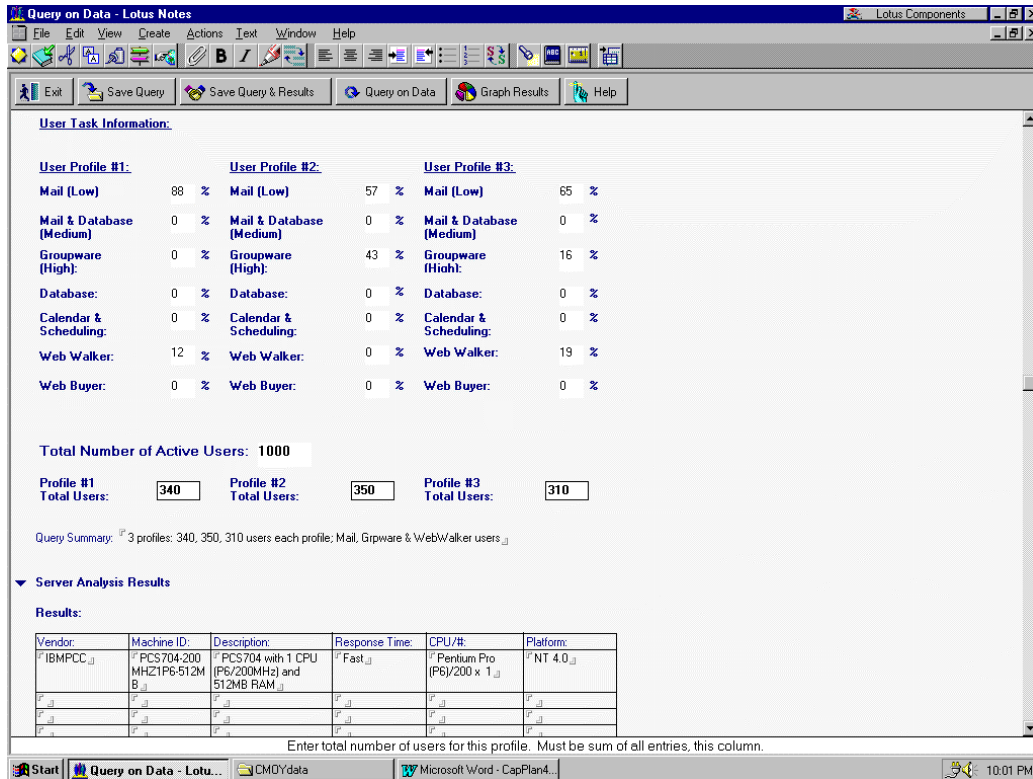


Figure 13

---

## Example Capacity Planning Exercise

---

This section applies the simple queries and user profiles discussed in previous sections to a more realistic capacity planning exercise.

The exercise is performed on behalf of a hypothetical organization; a nationwide service company with several branch offices, each of which has basically identical staffing composition. The employees in each branch office can be classified as either “principals” or “support staff.” The total staff at each branch office is 150 employees: 50 principals and 100 support staff. All employees will be supplied with their own Notes client.

The average Domino-related work activity of a principal is represented as 80% Mail and 20% Calendaring & Scheduling. This activity level is using Profile 1.

The average Domino-related work activity of a support staff person is represented as 30% Mail, 30% MailDB and 40% Calendaring & Scheduling. Support staff can fall into one of two or three different sub-categories.

In a typical 150-person branch office, the support staff may consist of three types of employees:

- 30 secretarial workers who compose mail and screen mail for one or more principals (represented as Mail activity).
- 30 researchers who work with shared discussion databases and communicate through Mail in support of the principals and their clients (represented as MailDB activity).
- 40 personal assistants who schedule appointments and manage the calendars for one or more principals (represented as Calendaring & Scheduling activity).

These three types of employees are grouped into one profile because there aren’t enough profiles available to describe each type of support staff separately.

In the query form, the default value for % Query Tolerance is 10%. Thus, for a query with 100 Mail users, any data point with 90 (100 minus 10%) or more users will satisfy this query component. You may wish to avoid higher values for this parameter, because it creates an effective minimum user requirement that can be difficult to keep track of in a complicated query specifying multiple NotesBench scripts.

As recommended above, initial queries specify a “slow” Response Time parameter. This means that Server.Planner will accept data points regardless of whether their Average Probe Response Time was fast, medium or slow. The query result will show the actual Average Probe Response Time for each accepted data point; evaluators can then apply more strict Response Time criteria if multiple choices are returned.

If Response Time is “slow,” the value of % Response in Range is immaterial. You may wish to leave this parameter set at 50%; thus, if you specify a fast Response Time, only half of the Probe Response Time samples need be in the “fast” category.

With this user profile and parameter specification, our hypothetical analysts generated seven queries (shown in Figure 14) for different numbers of total supported users using the IBM Personal Systems Group Server.Planner database downloaded from the NotesBench Consortium Website in September, 1998.

<i>Query#</i>	<i>Total number of users supported / multiple of the basic unit</i>	<i># users in Profile1 (80% Mail, 20% C&amp;S)</i>	<i>#users in Profile2 (30% Mail, 30% MailDB, 40% C&amp;S)</i>	<i>Systems that satisfy query</i>
1	150 (the basic unit)	50	100	704, 330, 325, NF3500
2	900 / 6	300	600	704, 330, 325
3	1050 / 7	350	700	704, 330, 325, NF7000
4	1200 / 8	400	800	704, 330, NF7000
5	1500 / 10	500	1,000	704, 330, NF7000
6	1800 / 12	600	1,200	704, NF7000
7	1950 / 13	650	1,300	704, NF5500, NF7000

*Figure 14*

The purpose of Query #1 is to find a system that could meet the basic requirements of a typical branch office. Based on price, the analysts might choose the Netfinity 3500. Although the configuration upon which the data is based has only one CPU and two drives, it can be expanded to a two-processor SMP server with a RAID array, to support a higher number of users. This high level of expandability provides for future growth of applications usage at each branch office.

For a regional office with a larger number of employees but the same basic employee composition, other queries in Figure 14 are relevant. The total number of supported users in each query is a multiple of that used in Query #1. Query #2 shows that three servers support an office six times the size of the “basic unit,” while Query #3 shows that PC Servers 704, 330, 325, and NF7000 support an office seven times the base size. As the total number of supported users rises, fewer and fewer server configurations meet requirements.

### **A different interpretation of the data**

Depending on requirements, the customer could interpret the data in a completely different way. For example, each of the server configurations could be upgraded to support additional Domino services for users connected via a Wide Area Network. One of the servers returned by Query #2 through Query #7 could be installed in a regional office to supports users connecting through the WAN from multiple remote branch offices. Here, analysts would need to consider network costs, server administration and any additional support costs.

## **A different way to represent user activity**

The hypothetical analysts could choose to represent the work activity of the support staff in a different, more complex manner. For example:

- Group 1: 40 people doing Mail and some Calendaring & Scheduling
- Group 2: 40 people doing MailDB and some Calendaring & Scheduling
- Group 3: 20 people doing only Calendaring & Scheduling

The amount of time the group spends on each activity would be expressed in terms of numbers of people working full-time on a single activity. Perhaps members of Group 1 spend 80% of their time doing Mail and 20% doing Calendar & Scheduling; this equates to 32 full-time Mail users and 8 full-time Calendaring & Scheduling users. By converting the activity of each group into equivalent number of full-time people performing each type of work, and then adding up the number of equivalent “actual” people performing each type of work, one could describe a more complicated user composition in a single profile.

In fact, this is conceptually how Server.Planner reduces a complex query to a more manageable one before it searches the vendor database for matches. It can then search the database using the fraction of the total users represented by a specific script.



---

## **Future directions**

---

The IBM Personal Systems Group plans to release more NotesBench data for use with Server.Planner on an ongoing basis. IBM customers can rest assured that any NotesBench data sets we release will continue to produce valid results with Server.Planner. We also plan to include supplemental performance information with our dataset; i.e., information not necessarily utilized by Server.Planner.

The effectiveness of Server.Planner depends not only on its functional capabilities, but also on the methodology hardware vendors use in generating their datasets. We will continue to refine our methodology to improve the accuracy of the results obtained with Server.Planner and reflect the performance of future releases of the Domino Server.