May–June 2001

Retek Planning Application Scalability Benchmark

Prepared by John Morris, Retek
Ralf Schmidt-Dannert, IBM

Contents

List of figures

List of tables

Executive summary

This document presents the results of a benchmark run by Retek and IBM for a retail customer. The objective of the benchmark was to provide an early indication of the performance of the Retek Predictive Application Server (RPAS) running on IBM hardware with the Retek TopPlan planning application. Planning parameters included:

- The expected volume of data
- The expected number of plannable items
- The expected number of users
- A metric set representative of an actual customer-planning environment

The benchmark was not intended to determine the highest application and system throughput achievable with “unlimited” hardware resources and extensive OS and/or application tuning. In fact, Retek performed only limited tuning, and throughput was limited by the available I/O subsystem.

Retek executed the benchmark using a generic TopPlan implementation built on Retek RPAS Version 9.3.5 and a customer-specific planning environment. IBM configured the hardware in preparation for the benchmark and provided technical assistance in configuring the performance-monitoring framework.

The benchmark results demonstrate that Retek's suite of planning applications running on IBM hardware delivers excellent performance under heavy load for a variety of planning tasks. Key performance statistics are presented in Table 1. These statistics were gathered under heavy load, with interactive runs performed in a planning environment that had 600,000 plannable items and 270 users.

Table 1. Key performance statistics

Performance statistics	Time
Mean calculation time for an edit to sales at the month level spreading to week (includes updates of dependent metrics and displayed aggregates).	14 sec.
Mean time to expand product and calendar dimensions to display aggregate values.	7 sec.
Mean time to switch products using the VCR control.	2 sec.
Mean time to switch between tabs.	1 sec.
Duration of batch process.	90 min.

This paper describes the benchmark scope, rationale, objectives, test configuration, test setup, results and conclusions.

Scope

The benchmark encompassed the interactive (in-season and pre-season) and batch activities of a top-down, bottom-up planning application deployed across a large retail customer—including all brands and all companies. The benchmark was limited to assessing the performance of the application server on the selected IBM hardware. Assessments of the RPAS client GUI and customer systems that would complete a planning solution implementation were beyond the scope of the benchmark.

Retek TopPlan

TopPlan is an easy-to-use, Web-based application that provides top-down, bottom-up and middle-out functionality with the ability to fully reconcile and approve plans. TopPlan employs an industry-standard process and provides sufficient scalability to support planning at levels of detail appropriate for any business.

Built on a powerful predictive engine, TopPlan uses integrated demand forecasting to provide a quick and accurate view of customer demand with minimum user intervention. Exception management functionality identifies affected areas of a plan that may otherwise go unnoticed when managing large amounts of data.

Retek TopPlan is based on the advanced RPAS server that supports scalability to levels as low as customer/store/SKU/week. The technology supports both Web deployment and traditional client-server deployment environments.

Hardware environment

Retek developed and executed two separate benchmarks for two types of workloads: interactive and batch. Both interactive and batch tests were performed on the hardware configuration shown in Figure 1. The hardware included an IBM @server pSeries 680 platform and two injector systems that simulated user interaction during the interactive runs. Table 2 lists the server hardware. Table 3 lists the hardware for the two injector systems. All runs were performed using the same hardware configuration and operating system settings.

Figure 1. Hardware configuration

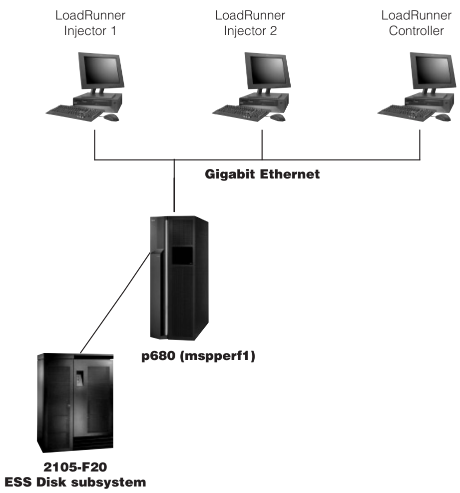


Table 2. Benchmark server hardware

Server hardware	
Vendor	IBM
Platform	p680
OS Version	AIX® 4.3.3, ML8
CPU	24-way 600MHz RS64 IV
Storage	IBM TotalStorage™ Enterprise Storage Server™, 2105 F20, 16GB cache, 1.7TB disk space (36GB drives); connected directly with eight fiber channels to the p680 server
Memory	64GB

To derive optimal performance from the available IBM Total Storage Enterprise Storage Server (ESS), Retek created two 50GB LUNs on each of the 8-disk ranks in the ESS. These LUNs were mapped in AIX through SDD (Subsystem Device Driver) to vpath0 – vpath15. Retek created the volume group vg01 (400GB) using vpath0 – vpath7 and vg02 (400GB) using vpath8 – vpath15. The actual benchmark data was stored in a logical volume created in vg01. A backup of the benchmark data was kept on a second logical volume created for that purpose in vg02. See Appendix B for more information about ESS.

Table 3. Injector hardware

Injector hardware	Controller	Injector 1	Injector 2
Platform	x86	x86	x86
OS Version	Microsoft® Windows® 2000	Microsoft Windows 2000	Microsoft Windows NT® 4.1
CPU	400Mhz Pentium®	2-way 850Mhz Pentium	2-way 850Mhz Pentium
Memory	128MB	1GB	1GB

Hardware and operating system monitoring employed IBM nmon, svmon, iostat and vmstat monitoring tools.

Application configuration

Retek performed the benchmark using a generic TopPlan configuration with the core implementation parameters shown in Table 4. The parameters shown represent the planning environment for a large retailer.

Two test scenarios were developed: one interactive and one batch. The interactive scenario is representative of a high-activity planning period such as that found at the beginning of planning season when the start times of a large number of planners tend to be synchronized. The batch scenario is representative of the activities typically performed on a weekly basis to load actualized data and product hierarchies. The weekly batch activities are a superset of activities—such as additions and reclassifications—that are performed on a nightly basis to maintain product hierarchy. Table 5 shows the distribution of active and inactive users by planning level. Table 6 shows the plan size by planning level. Table 7 shows the distribution of planning activities.

Table 4. Core implementation parameters

Core implementation parameters	Value
User Count	270
Style-Color Count	600,000
Number of Planning Tiers	3 (planner, manager, executive)
Data Feed Level	Style-Color/Week
Number of Data Feeds	32 (31 metrics plus 1 product hierarchy)
Weekly Data Feed Volume	1,000,000 records
Number of Weekly Product Additions	3,000
Number of Weekly Product Reclassifications	1,000

Table 5. Virtual user distribution in interactive scenario planning

Virtual user distribution						
Planning Level	Active			Inactive		
	Planner	Manager	Executive	Planner	Manager	Executive
Top-Down	0	24	10	0	26	10
Bottom-Up	96	0	0	104	0	0

Table 6. Plan size

Plan size

User Class	Product Level	Plannable Items	Calendar Level	Plannable Weeks
Planner	Style-Color	140	Week	13
Manager	Class	25	Week	13
Executive	Division	5	Week	13

Table 7. Planning scenario activities

Planning scenario activities

Group	Activity	Description	Percent of Users Engaged in Activity
Plan Operations	Create Plan	User creates a plan using the interactive wizard procedure.	2.5
	Open and Close Plan	User opens and closes a plan.	5.0
	Save Plan	User saves an open plan.	5.0
	Commit Plan	User commits a plan to the data cache.	2.5
	Refresh One Metric	User updates a single metric in a plan.	5.0
	Refresh All Metrics	User updates all metrics in a plan.	2.5
Navigation	Switch Between Tabs	User switches from tab to tab in an open plan.	7.5
	Scroll	User scrolls through a worksheet in an open plan.	7.5
	Pivot	User pivots a worksheet in an open plan.	5.0
	Locking	User locks and unlocks data points on a worksheet in an open plan.	5.0
	Collapse and Expand	User collapses and expands rollups in a worksheet in an open plan.	7.5
	VCR Control Expand	User moves through the products in a plan using the VCR control.	5.0
Calculation	Calculate	User edits a plan and initiates calculations.	40.0

Benchmark execution

Retek executed five interactive graded runs using Mercury Interactive LoadRunner. LoadRunner was augmented with a DLL that simulated the communication backend of the RPAS client. A parameterized LoadRunner script was employed to load the DLL, logon to the application server (via the DLL) and execute commands (again via the DLL). User-specific parameters determined which command set was to be executed.

To accommodate the large number of users, each interactive run was divided into three phases. In the first phase, all simulated users logged in. In the second phase, 124 of the 130 total active users opened plans as specified by their command set and then proceeded to a rendezvous. All other users proceeded immediately to the rendezvous point. When all users had reached the rendezvous point, they were released to perform the tasks dictated by the command set. Prior to each run, the interactive test environment was restored to its original state. A hardware-monitoring program was launched before completion of the first phase to collect system performance statistics during the course of the run. At the end of the run, the hardware-monitoring program was terminated.

Retek also performed five graded batch benchmark runs. Prior to each batch run, the batch benchmark environment was restored to its original state. The invocation and termination of the hardware-monitoring program was integrated with the batch execution program.

Results

This section presents the benchmark results, which are divided into two categories: interactive and batch. The results in the interactive category are further divided into planner, manager and executive response times, and system performance statistics. The results in the batch category are further divided into batch run times and system performance statistics. Retek performed more than 20 interactive runs. The baseline graded sequence was derived from run numbers 12, 13, 14, 15 and 16. Retek also performed five batch runs.

Interactive performance statistics

This section describes both the planning response times and the system performance statistics for the interactive runs.

Planning activity response times

This section describes the planning activity response times for the interactive runs. Table 8, Table 9 and Table 10 present consolidated views of planner response times for runs 12 to 16 for each planner tier. Retek established target values for the response times for each of the minor planning activities. These target values, shown in the tables, represent Retek's estimation of the response times necessary to maintain high user productivity.

Table 8. Consolidated planner response times for interactive sequences 12–16

Consolidated planner response times				Time (seconds)						
Group	Major Activity	Minor Activity	Users	Min	Avg	Max	75th Percentile	90th Percentile	Target	Cycles
Plan Operations	Commit Plan	Commit Plan	2	21	31	196	36	41	40	100
	Create Plan	Create Plan	2	172	391	1,260	210	223	180	10
	Open and Close Plan	Close Plan	5	1	2	4	3	3	4	60
		Open Plan		34	53	100	60	67	45	60
	Refresh All Metrics	Refresh All Metrics	3	89	231	649	192	230	120	12
	Refresh One Metric	Refresh One Metric	5	28	58	169	77	92	30	125
	Save Plan	Save Plan	5	8	27	81	39	48	45	210
Navigation	Expand and Collapse	Collapse Calendar	7	1	5	12	6	8	4	350
		Collapse Product		1	4	10	5	7		350
		Expand Product and Calendar		2	7	17	9	12		350
		First Collapse Calendar		5	6	8	7	7		7
		First Collapse Product		5	6	7	6	7		7
		First Collapse Product and Calendar		29	30	32	31	31		7
	Locking	Locking	5	1	2	13	2	2	4	1,500
	Pivoting	Pivoting	5	1	2	11	3	3	4	250
	Scrolling	Scrolling	7	1	2	11	3	3	4	350
	Switch Tabs	First Tab Visit	7	8	12	25	12	15	4	63
		Switch Tabs		0	1	4	1	1	4	3,850
	VCR Control	VCR Control	5	1	2	13	3	3	4	250
Calculate	Calculate	Calculate	38	6	14	27	18	21	20	1,950

Notes:

1. No significant differences were observed between in-season and pre-season response times.
2. Response times less than 0.5 seconds are reported as 0.
3. First-time expand numbers are higher than subsequent expansions because aggregates are calculated on an as-needed basis. As a result, they are not calculated until they need to be displayed.
4. First-time switch tab times are higher than subsequent switch tab times because at the first tab visit the server must prepare aggregates for display and send the tab display information to the client.

Table 9. Consolidated manager response times for interactive sequences 12–16

Consolidated manager response times				Time (seconds)						
Group	Major Activity	Minor Activity	Users	Min	Avg	Max	75th Percentile	90th Percentile	Target	Cycles
Plan Operations	Create Plan	Create Plan	1	209	420	1,193	218	265	180	5
	Open and Close Plan	Close Plan	1	1	2	4	3	3	4	15
		Open Plan	1	37	57	106	62	68	40	15
	Refresh All Metrics	Refresh All Metrics		101	231	499	226	256	120	6
	Refresh One Metric	Refresh One Metric	1	39	67	174	88	106	30	25
Navigation	Save Plan	Save Plan	1	8	26	68	38	47	45	45
	Expand and Collapse	Collapse Calendar		1	4	10	6	8		100
		Collapse Product		1	4	9	5	6		100
		Expand Calendar		1	4	9	6	7		100
		Expand Product		2	6	15	8	11		100
		First Collapse Calendar	2	7	8	8	7	7	4	2
		First Collapse Product		7	7	8	7	7		2
		First Expand Calendar		25	25	25	25	25		2
		First Expand Product		11	12	13	11	11		2
	Locking	Locking	1	1	2	12	2	2	4	300
	Pivoting	Pivoting	2	1	3	10	3	4	4	65
	Scrolling	Scrolling	2	1	3	10	3	3	4	100
	Switch Tabs	First Tab Visit	2	8	12	23	12	13	4	18
	Switch Tabs			0	1	4	1	1	4	1,100
	VCR Control	VCR Control	1	1	3	9	3	4	4	50
Calculate	Calculate	Calculate	10	6	15	32	19	21	20	500

Note:
1. No significant differences were observed between in-season and pre-season response times.
2. Response times less than 0.5 seconds are reported as 0.
3. First-time expand numbers are higher than subsequent expansions because aggregates are calculated on an as-needed basis. As a result, they are not calculated until they need to be displayed.
4. First-time switch tab times are higher than subsequent switch tab times because at the first tab visit the server must prepare aggregates for display and send the tab display information to the client.

Table 10. Consolidated executive response times for interactive sequences 12–16

Consolidated executive response times				Time (seconds)						
Group	Major Activity	Minor Activity	Users	Min	Avg	Max	75th Percentile	90th Percentile	Target	Cycles
Plan Operations	Open & Close Plan	Close Plan	1	2	2	4	3	3	4	15
		Open Plan		26	50	143	52	60	45	15
	Refresh One Metric	Refresh One Metric	1	15	24	101	23	25	30	25
Navigation	Expand & Collapse	Collapse Calendar	1	1	4	9	5	7	4	50
		Collapse Product		1	4	9	5	6		50
		Expand Calendar		1	4	10	6	8		50
		Expand Product		1	5	12	7	9		50
		First Collapse Calendar		6	6	6	6	6		1
		First Collapse Product		7	7	7	7	7		1
		First Expand Calendar		26	26	26	26	26		1
		First Expand Product		10	10	10	10	10		1
	Scrolling	Scrolling	1	1	2	8	3	3	4	50
	Switch Tabs	First Tab Visit	1	10	14	23	10	10	4	3
		Switch Tabs		0	1	2	1	2		300
	VCR Control	VCR Control	1	1	2	8	3	3	4	50
Calculate	Calculate	Calculate	4	5	12	22	15	17	20	200

Notes:

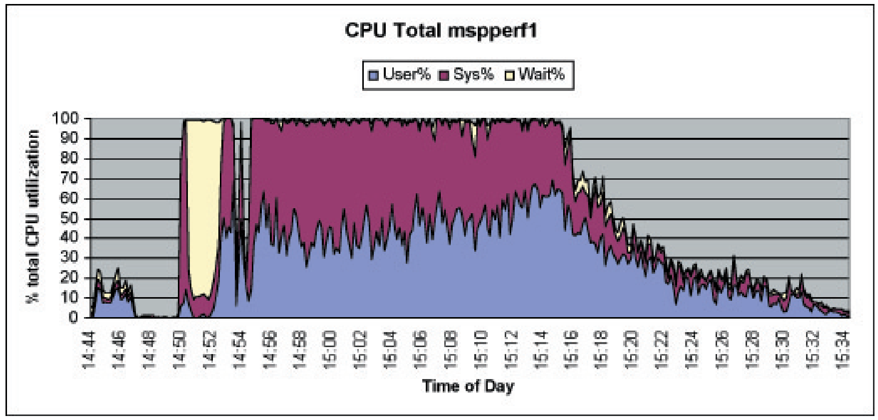
1. In generic TopPlan, executive-level planning does not establish a distinction between in-season and pre-season.
2. Response times less than 0.5 seconds are reported as 0.
3. First-time expand numbers are higher than subsequent expansions because aggregates are calculated on an as-needed basis. As a result, they are not calculated until they need to be displayed.
4. First-time switch tab times are higher than subsequent switch tab times because at the first tab visit the server must prepare aggregates for display and send the tab display information to the client.

System performance statistics

This section describes the system performance statistics for the interactive runs. The interactive processing profile has three distinct phases. During the first phase all virtual users are activated through the LoadRunner control station, about 5–10 at a time, and logon to the TopPlan application server. The second phase is characterized by a very high I/O wait percentage. During this period, virtual users open workbooks. Approximately 130 users open their specific workbooks in the four to five minute period that constitutes this second phase. Prior to the third phase, all virtual users are suspended at a rendezvous point. When all virtual users have reached this point, they are released en masse and proceed with their planning task as defined in the different interactive user roles.

Figure 2, Figure 3, Figure 4 and Figure 5 show the CPU utilization, process load, physical memory used and disk total during interactive run 15. Table 11 shows the summarized adapter I/O statistics for interactive run 15. Logging of the trace data was done over the SCSI controllers. Table 12 shows the total system I/O statistics for interactive run 15.

Figure 2. CPU utilization for interactive sequence 15



Note: Interval length = 10 seconds.

Figure 3. Process load for interactive sequence 15

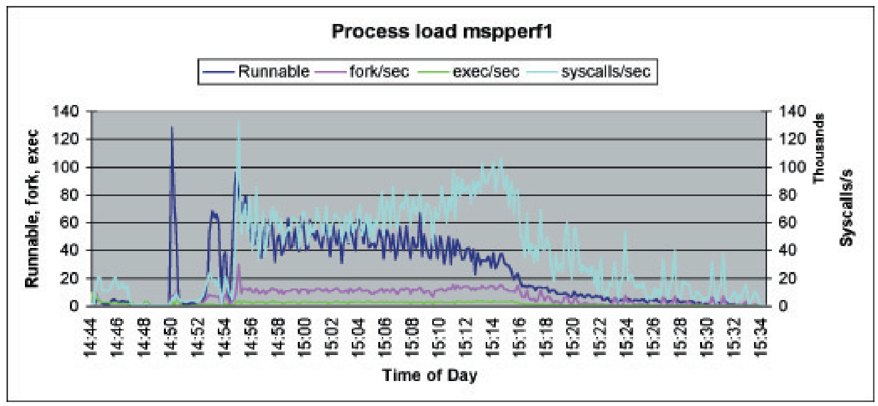


Figure 4. Physical memory used for interactive sequence 15

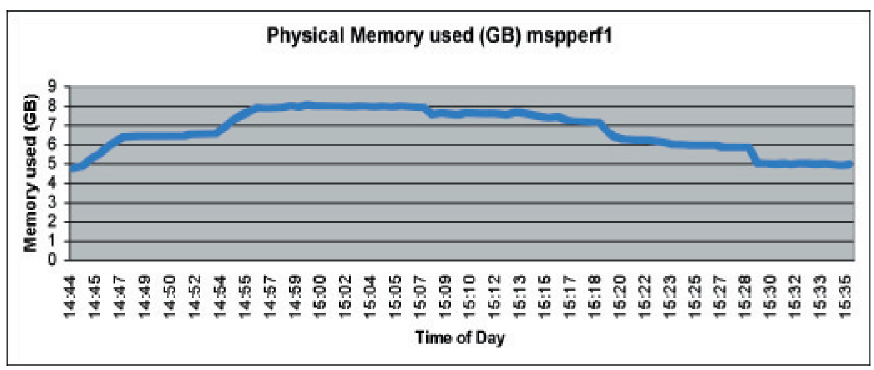
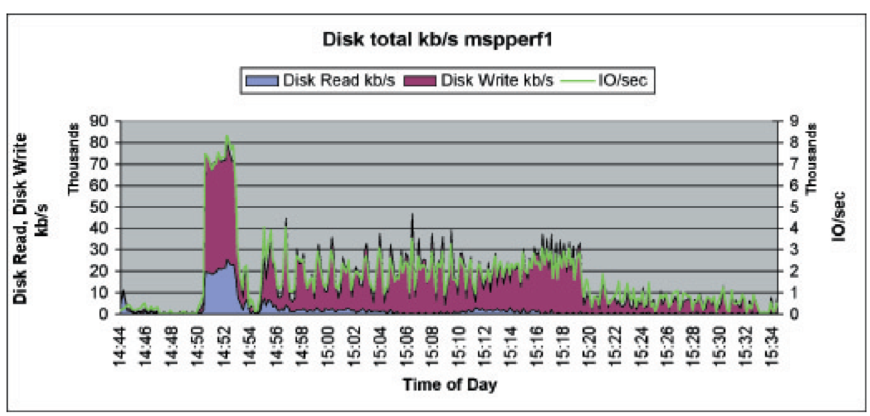


Figure 5. Disk total for interactive sequence 15



Batch performance statistics

This section describes both the batch times and the system performance statistics for the batch runs.

Batch times

Table 13 presents the application build statistics, that is, the times required to perform the initial hierarchy and data load on the configured TopPlan application. Table 14 presents the batch performance statistics for each of the five graded batch runs. Table 15 presents an estimate of batch plan build time. Large plan builds (140 plans) were performed to simulate the batch build process. The batch build time presented in Table 15 represents a likely upper bound on the batch build time for 140 plans.

Table 11. Adapter I/O statistics for interactive sequence 15

Adapter I/O statistics for interactive sequence 15

Adapter	Avg Kbps	Max Kbps	Avg tps	Max tps	Total KB read	Total KB written
fscsi0	1,925	9,909	201	1,009	684,218	5,192,796
fscsi1	1,922	9,602	201	1,018	74,977	5,192,056
fscsi2	1,920	9,764	201	1,039	672,346	5,186,548
fscsi3	1,919	9,902	201	1,060	661,091	5,196,236
fscsi4	1,919	9,668	201	1,032	671,458	5,186,108
fscsi5	1,915	9,368	201	992	668,290	5,180,088
fscsi6	1,916	9,414	200	1,017	670,084	5,180,184
fscsi7	1,923	9,388	201	1,010	674,948	5,195,376

Table 12. Total system I/O statistics for interactive sequence 15

Total system I/O statistics for interactive sequence 15

Max. transactions/second during an interval (R/W)	8,114
Avg. transactions/second during an interval (R/W)	1,631
Total number of Kbytes read	5,377,412
Total number of Kbytes written	41,509,392
Read/Write Ratio	0.13

Note: Interval length = 10 seconds.

Table 13. Application build statistics

Application build statistics

Operation	Duration (hours)
Stock Ledger Full Product Master Update	4.0
Stock Ledger Initial History Load	16.0
User Partition Full Product Master Update	1.0
User Partition Initial History Load	0.5

Table 14. Batch performance statistics

Batch performance statistics

	Time (minutes)						
	Batch Sequence No.					Mean	Std. Dev
Operation	1	2	3	4	5		
Batch Process Duration	71.77	82.98	82.33	82.77	85.17	81.00	5.28
Stock Ledger Delta Product Hierarchy Update	7.63	7.65	8.33	8.32	8.15	8.02	0.35
Stock Ledger Data Load and Metric Re-Shape	12.781	2.23	12.78	13.57	13.32	12.94	0.52
Stock Ledger Metric Re-Sync	15.88	18.03	17.33	18.22	18.25	17.54	1.00
User Partition Delta Product Hierarchy Update and Metric Re-Shape	20.10	29.00	28.87	28.42	30.95	27.47	4.23
User Partition Data Load and Metric Re-Sync	28.08	28.25	27.73	27.77	27.78	27.92	0.23

Table 15. Batch plan build

Batch plan build

Operation	Duration (minutes)
Batch Build of 140 Plans	~45

Operation Duration (minutes) Batch Build of 140 Plans~45

System performance statistics

The batch processing profile is characterized by two widely spaced humps in CPU utilization and I/O rate. During the first hump, the batch process updates the product hierarchy in stock ledger and user partitions. Following this update, both data arrays in both the user partition and stock ledger are reshaped. When reshaping is completed in the stock ledger, data load begins. All of these tasks are carried out in parallel. During the low CPU utilization period, stock ledger data is resynchronized. This is a serial process. In the second hump, the batch process loads data into the user partitions. When this data load is complete, the data in the user partitions is resynchronized. The procedures executed during the second hump are performed in parallel.

Table 16. Adapter I/O statistics for batch sequence 4

Adapter I/O statistics

Adapter	Avg Kbps	Max Kbps	Avg tps	Max tps	Total KB read	Total KB written
fscsi0	2,972	11,725	151	745	2,875,881	12,110,860
fscsi1	2,979	11,368	151	814	2,926,901	2,092,052
fscsi2	2,974	11,532	151	719	2,897,824	12,098,124
fscsi3	2,960	11,486	151	704	2,883,699	12,043,168
fscsi4	2,971	11,368	151	751	2,877,271	12,103,528
fscsi5	2,996	11,337	152	715	2,914,481	12,194,844
fscsi6	2,973	11,198	152	785	2,910,899	12,079,388
fscsi7	2,960	11,858	151	736	2,895,885	12,028,440

Table 17. Total system I/O statistics for batch sequence 4

Total system I/O statistics for batch sequence 4	
Max. transactions/second during an interval (R/W)	6,024
Avg. transactions/second during an interval (R/W)	1,224
Total number of Kbytes read	23,182,870
Total number of Kbytes written	96,750,404
Read/Write Ratio	0.24

Note: Interval length = 10 seconds.

Figure 6, Figure 7, Figure 8 and Figure 9 show the CPU utilization, process load, physical memory used and disk subsystem load during batch run 4. Table 16 shows the adapter I/O statistics for batch run 4. Table 17 shows the total system I/O statistics for batch run 4.

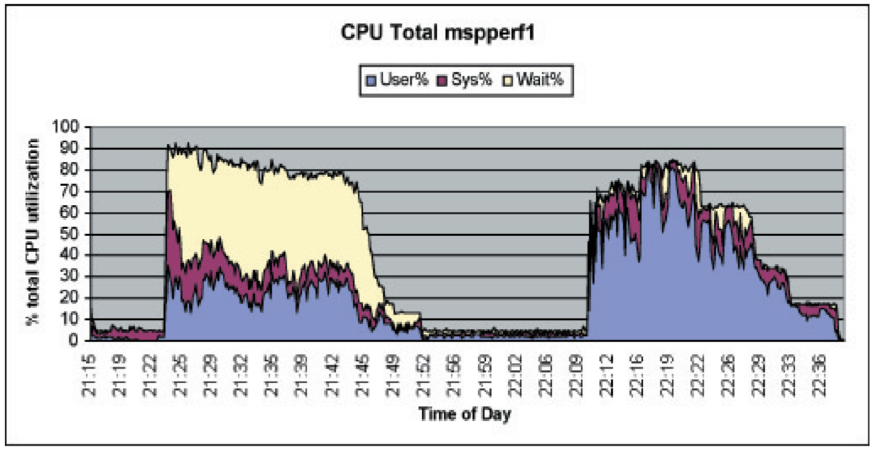


Figure 6. CPU utilization for batch sequence 4

Note: Interval length = 10 seconds.

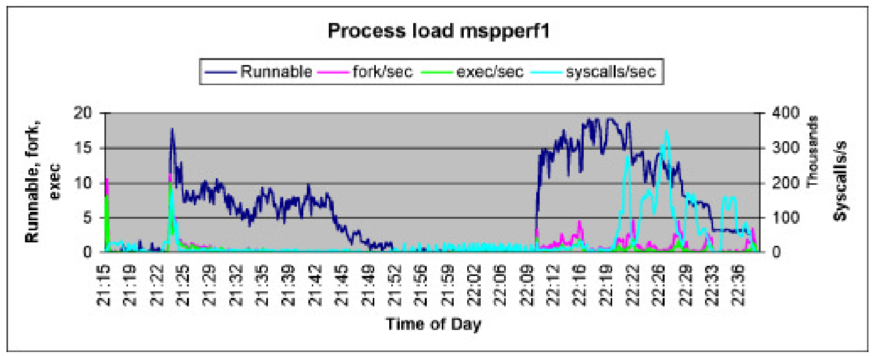


Figure 7. Process load for batch sequence 4

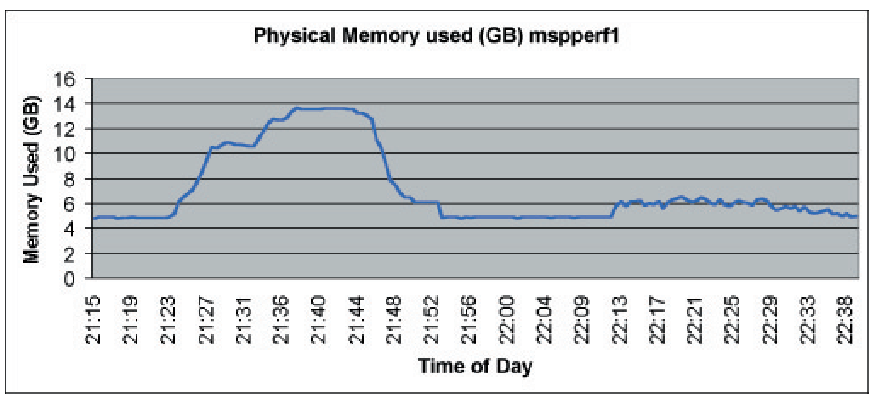


Figure 8. Physical memory used for batch sequence 4

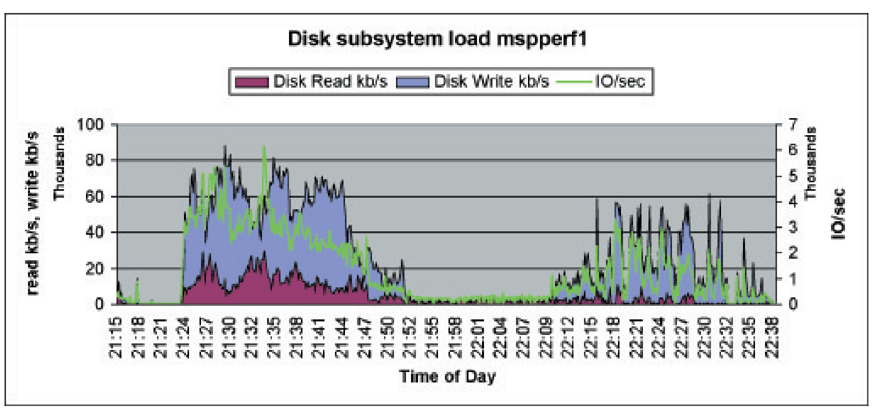


Figure 9. Disk subsystem load for batch sequence 4

Analysis of results

The benchmark results demonstrate that Retek's suite of planning applications running on IBM hardware delivers excellent performance overall for a variety of planning tasks, even under heavy load (600,000 plannable items and 270 users). There were, however, some areas in which the TopPlan application missed the specified target values by a significant amount. These areas are discussed in the remainder of this section. For each area, the suspected causes of the anomalies are examined and methods are suggested to mitigate the effects of the anomalies. The discussion is divided into two sections: interactive and batch.

Analysis of interactive results

–Long wait for I/O

The extensive I/O wait (80 to 90 percent) observed in the first high-load peak is an artifact of the benchmark setup. 124 users open their workbook at about the same time. This operation first creates a backup of the current user specific workbook (40MB–60MB) before actually opening the workbook. In an actual production environment, this workload would be distributed over a longer time period and would result in significantly less peak load on the I/O subsystem.

–System time

The planning application exhibits unusually high system time consumption relative to user time during interactive workload. This type of behavior is usually indicative of application inefficiencies. The high system time contributes to the high CPU utilization observed during interactive runs. As a result, a reduction in system time would probably improve overall performance. An initial detailed performance analysis was performed that indicated several potential areas for future application tuning. Retek is currently addressing these areas and expects to achieve a significant reduction in the ratio of system time to user time.

–Plan create and refresh times

As Table 18 shows, the plan create and refresh times (and to a lesser extent the plan open times), significantly missed the target times. Retek believes that these times need to be improved to maintain a high level of user productivity.

In all runs, the first create plan cycle required the longest amount of time. During this period the system load was at peak levels, with all designated virtual users actively performing their tasks. The generated load typically dropped off rapidly during the second and following create plan runs, because more and more of the other virtual users already finished their defined work scenarios. For this reason, the first cycle is more representative of the duration of the create plan task under high machine load than the 90th percentile. A similar approach was used to estimate the duration of the refresh all metrics task.

Table 18. Benchmark target - Actual mismatches

Performance metric	Target (sec)	Actual (sec)
Refresh One Metric	30	100
Refresh All Metrics	120	400
Create Plan	180	1100

Retek is currently implementing modifications to the RPAS architecture to improve plan create and refresh response times. The modifications are scheduled for inclusion in the next major RPAS release (planned for release in mid-2002). In addition, the use of more powerful IBM hardware, such as the IBM @server pSeries 690 platform with its superior floating point throughput, would significantly reduce the response times for calculation-intensive activities such as metric refresh and workbook creation.

Customers can mitigate the impact of the longer than targeted plan create and refresh times in a variety of ways. One way is to implement administration practices that minimize the effects. For example, a customer can implement an administration process that minimizes the number of plan creations and refreshes that users need to perform. Another way is for the customer to employ nightly batch plan create and refresh processes to reduce the time users spend managing their plans. Finally, in view of the strong correlation between metric count and performance, a customer might also implement lighter-weight plans.

Analysis of batch results

The benchmark exhibited an excellent batch run time of approximately 82 minutes. Improving CPU utilization by distributing the planning data across more planning domains and having at least one domain per CPU can make this fast batch run time even faster. (The benchmark system had only 20 domains defined with 24 CPUs actually in the system.)

In addition, increasing the I/O bandwidth would reduce batch run times, especially in the first of the two high CPU load phases during the batch test. In that phase, a separate process for each planning domain is spawned to update the product hierarchy and reshape user partition metrics.

Conclusions

As shown in Table 19, the benchmark results demonstrate that Retek's suite of planning applications running on IBM hardware deliver excellent performance overall for a variety of planning tasks, even under heavy load (600,000 plannable items and 270 users).

Table 19. Planning application

Performance statistics

	Value
ValueMean calculation time for an edit to sales at the month level spreading to week including updates of dependent measures and displayed aggregates.	14 seconds
Mean time to expanding product and calendar dimensions to display aggregate values.	7 seconds
Mean time switching product using the VCR control.	2 seconds
Mean time to switch between tabs.	1 seconds
Batch process duration.	90 seconds

Retek intends to continually evolve the RPAS architecture. One of the major goals of this evolution is to boost performance on a continuing basis. (The next major RPAS release, scheduled for mid-year 2002, is expected to improve performance in a number of areas). This continually increasing speed, combined with the comprehensive functionality of TopPlan, ensures a highly productive planning environment now and into the future.

Appendix A: p680 server information

The p680 is a symmetric multiprocessing (SMP) server for high-end commercial performance. The p680 platform is ideally positioned for the future. It supports concurrent execution of 32-bit and 64-bit applications. It uses the new SOI technology enhanced RS64 IV processors and is offered in 6- , 12- , 18- , and 24-way configurations. In addition, the p680 server includes an enhanced L2 16MB cache and an increased memory size (96GB) for large database transactions.

For more information:

- http://www-1.**ibm.com**/servers/eserver/pseries/hardware/factsfeatures.pdf
- http://www-1.**ibm.com**/servers/eserver/pseries/hardware/whitepapers/p680_technology.pdf
- http://www-1.**ibm.com**/servers/eserver/pseries/hardware/whitepapers/p680_reliability.pdf

Table 20 presents an overview of the p680 server features.

Table 20. p680 server information

Server information	
Feature	Benefits
600MHz RS64 IV microprocessors	Greatly expand performance levels for SMP commercial applications Provide capacity to grow to 24 processors
Copper and SOI technology	Offers greater performance and reliability
16MB ECC L2 cache per processor	Provides increased performance and greater reliability
Up to 96GB ECC SDRAM memory	Allows faster performance and exploitation of 64-bit addressing as might be used with large database applications Provides growth options for significantly increased capacity
Chipkill™ Memory	Significantly lowers number of memory failures that cause system outages, thus increasing system availability Minimizes the potential for loss of business data
64-bit system architecture	Improves physical memory use for applications requiring faster access to large amounts of data
Up to 56 PCI adapter slots	Provide e-business growth options for significantly increased capacity Support many commonly used adapters for increased availability at lower costs
Hot-swappable disk drive bays	Provide greater system availability and smooth growth by allowing swapping or adding of disk drives without powering down the system
Redundant hot-plug power	Provide greater system availability since cooling fans or power supplies as well as cooling subsystems can be changed without interrupting the system
Built-in service processor	Continuously monitors system operations and takes preventive or corrective actions for quick problem resolution and high system availability Allows diagnostics and maintenance to be performed remotely
Dynamic Processor Deallocation	Automatically deallocates resources when impending processor failures are detected so applications can continue to run uninterrupted
Concurrent 32- and 64-bit	Allows customers to migrate to 64-bit applications at their own pace
Application support	Protects customers' 32-bit software investments while allowing them to participate in technology enhancements
Capacity Upgrade on Demand	Provides a cost-effective growth path in processing capacity Satisfies new, unanticipated user application demand rapidly
AIX clustering	Provides centralized management of multiple systems Provides ability to handle unexpected workload peaks by sharing resources Allows for more granular growth so user demands can be readily satisfied
AIX operating system	Maintains branded conformance to The Open Group's XPG4, UNIX® 95 and UNIX 98 specifications Provides an AIX binary-compatible environment that helps assure continuing application availability across AIX Version 4 releases Integrates IBM AIX Developer Kit, Java™ Technology Edition with IBM Just in Time (JIT) Compiler, allowing Java to load automatically in support of both server-side and client Java applications

Appendix B: ESS information

This section describes the IBM Enterprise Storage Server (ESS). The ESS is the third generation of the IBM Seascape® architecture for disk systems. It is a solution that provides the outboard intelligence required by SAN (Storage Area Network) solutions, offloading key functions from host servers and freeing up valuable processing power for applications. As a comprehensive SAN-based storage solution, the ESS provides considerable management flexibility to meet the fast-paced requirements of today's businesses.

Architecture overview

Figure 10. ESS functional topology

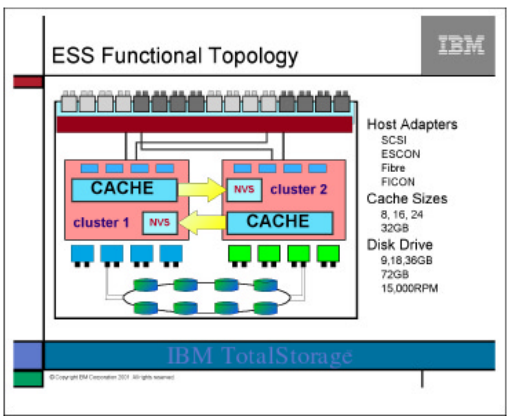


Figure 10 illustrates the major components of the ESS. There are 16 Host Adapters (HAs), which are the external interfaces. Each adapter supports two ports, either SCSI or ESCON®. Fibre Channel or FICON™ HA's support one channel per HA. Each HA connects to both Cluster Processor Complexes, so that either cluster can handle I/O from any host adapter.

There are two Cluster Processor Complexes in the ESS, each working independently. Each contains four RISC Symmetric Multi Processors (SMPs), with up to 16GB of cache (total combined up to 32GB), non-volatile cache (NVS) and four device adapters.

The design philosophy is that once data is in the ESS, it is to be fully protected. All host data is written to and from the cache using adaptive record, partial track, sequential or full track staging algorithms. In addition, all write data is automatically placed into two separate caches for integrity. The NVS cache has its own 7-day battery and is managed using a Least Recently Used (LRU) algorithm. On data writes, one copy of data is placed in one cluster and the second copy of the write data is placed in the NVS of the other cluster. In the event of either a planned or unplanned cluster outage, write data for the failed cluster is available on the other cluster and the remaining cluster takes over the functions in progress from the failed cluster.

Within each cluster, the Device Adapters (DAs) are used to connect disks to the Cluster Processor Complexes. DAs are always installed in pairs, one in each cluster. Disks arrays, or ranks (IBM 9, 18 or 36GB disks) are attached to the two DAs via 160 MB Serial Storage Architecture (SSA) loops. The ranks can be configured as RAID 5 or non-RAID arrays [Just a Bunch Of Disks (JBOD)].

The IBM ESS supports AS/400®, RS/6000®, NUMA-Q®, Netfinity®, Intel®, HP, Sun® and other host servers. For the complete list, see: www.storage.**ibm.com**/hardsoft/products/ess/supserver.htm The IBM ESS offers open systems users 24x7 availability with redundant design and non-disruptive maintenance. The StorWatch™ ESS Specialist enables centralized physical storage management, reducing and management complexity.

The IBM Subsystem Device Driver (SDD):

- Provides more than one path from the host to the ESS. A single LUN can appear as 2–16 LUNs.
- Offers host path failover.
- Supports load distribution across paths.
- Supports AIX and Windows NT, with other UNIX server types planned for the future.

The storage server standard for the new millennium
E-business is driving a data explosion, generating exponential growth in the need for immediately accessible, always available and highly functional storage capacity. In the increasingly e-centric world, information storage is becoming much like an electrical utility: You plug in, and you get what you need. Appliance-like intelligent storage served by its own network is required to provide “information as a utility.”

The IBM ESS is the ultimate SAN utility, providing the information “fuel” that runs the e-business “engine.” Extensive heterogeneous server connectivity makes the ESS a natural fit for server consolidation requirements. The ESS supports rapid universal access to vast quantities of data through many advanced functions and features, making it a workhorse for support of business intelligence and other business-critical applications. Because of its enterprise-wide support and management scope, the ESS is tailor-made to support consistent, efficient and effective enterprise resource planning. Using the ESS to address any or all of strategic and tactical business initiatives gives an organization the business advantage needed to survive and thrive in the e-world.

ESS highlights

- Ultra-high availability
- Massive scalability 420GB to 11.2TB
- High performance with I/O rates exceeding 45,700 per second
- SAN ready
- Advanced copy features
- Heterogeneous support for Windows NT, UNIX, AS/400 and S/390® data storage
- Business continuance options
- Investment protection
- Flexible IBM lease terms
- Price competitive
- Non-disruptive upgrade options
- RAID-5 or non-RAID
- Automatic I/O load balancing
- Synergy with IBM tape devices
- SCSI, Fibre Channel, ESCON, FICON
- Full hardware redundancy
- Web management tools

Appendix C: ESS benchmark configuration

The ESS in the benchmark was configured as shown in Figure 11, Figure 12 and Figure 13.

Figure 11. Shark F20 configuration

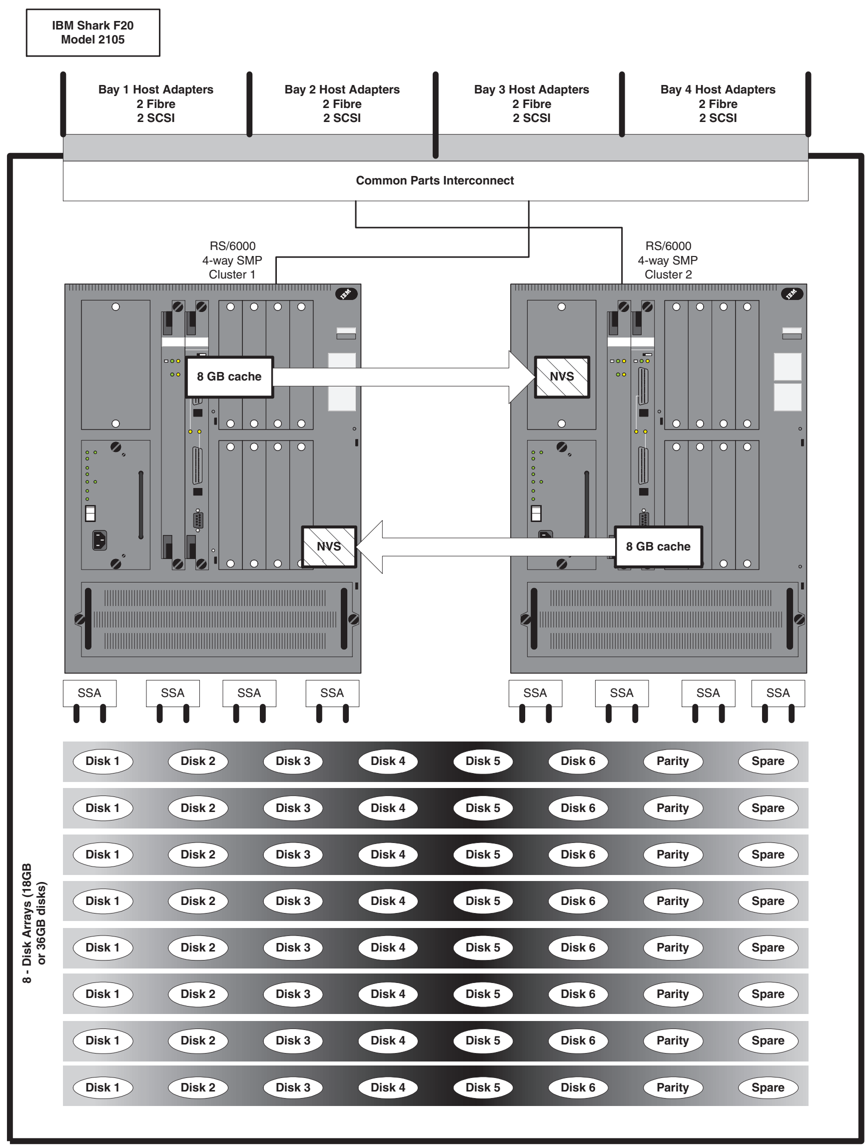


Figure 12. ESS - Server connection diagram

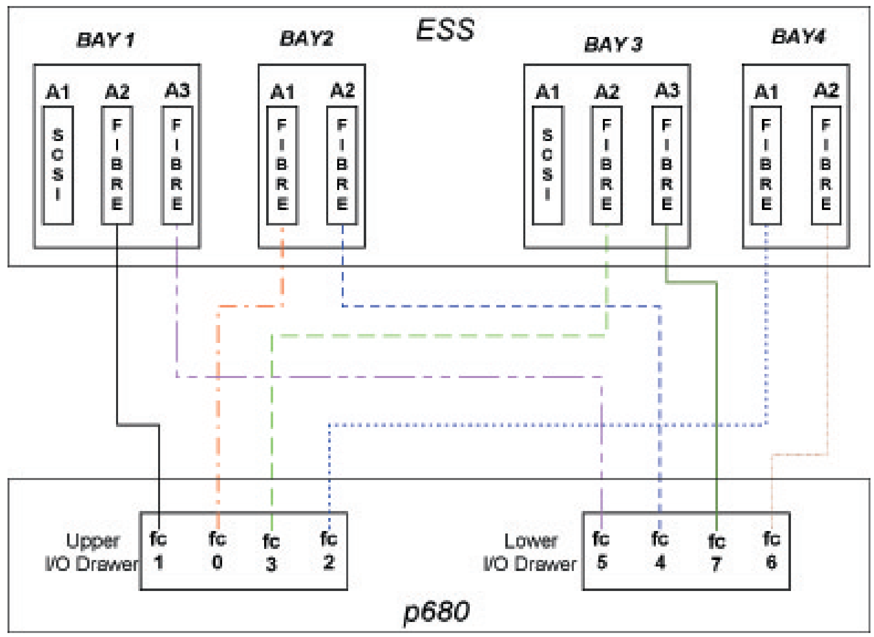
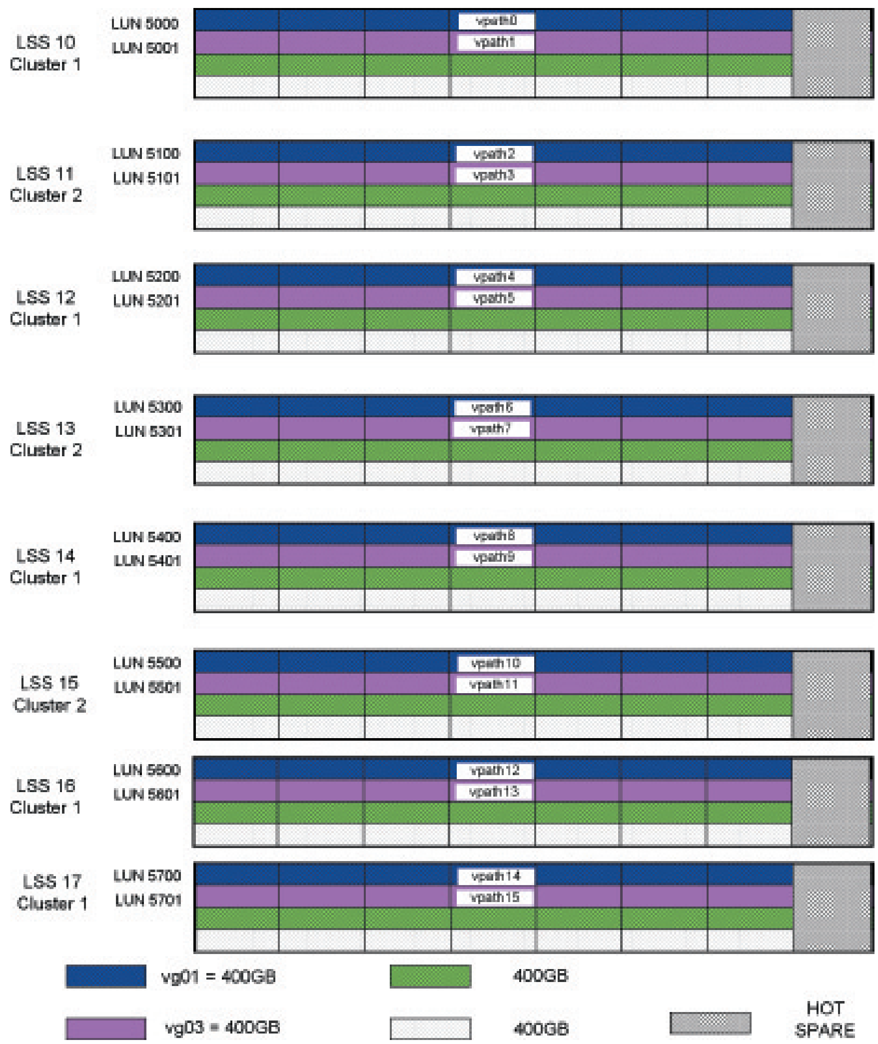


Figure 13. ESS - LSS, vpath and volume group layout



IBM Server Group
3039 Cornwallis Road
Research Triangle Park, NC 27709

Printed in the United States of America
1-02

This document contains performance information. Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the numbers stated here.

IBM and Retek are separate companies and each is responsible for its own products.

Printed on recycled paper containing 10% recovered post-consumer fiber.