MEGAPLEX MODELING

The Future of NonStop Demand

In the *Megaplex: an Odyssey of Innovation* and the follow-on *Road to the Megaplex* papers, we presented models based on changing the cost of application services from course-grained service levels to fined-grained service levels. In those papers, we present how the Megaplex model actually works and what benefits this model can bring to organizations in terms of improving their service levels while lowering costs. Here in the *Megaplex Modeling* paper, we explore these theories in more detail. We also take a hard look at other theories and how the service-oriented world is changing the nature of those theories.

KSU Costing Theory: The first theory we explore is utility pricing based on The Standish Group's KSU model (thousands of Standish Units). In our total cost of ownership (TCO) models, The Standish Group breaks down each of the application components (hardware, software, maintenance, manpower, and other costs) into dollars per KSU. Each of the cost units is stored in our CENTS database. We then normalize the application into transactions per KSU; followed by mapping the workload over a 24-hour period. The VirtualADVISOR application then multiplies the cost per item ($KSU) by transactions during both peak and off-peak hours. We then map that to create the yearly cost estimate, as shown in many tables in this report.

Black and White Availability Theory: For many years we presented the hierarchy of availability based on the 9's—the higher the 9's the more you paid and the higher the availability. On page four of the original Megaplex paper we presented a chart titled "Readiness Categories" that shows the percent of uptime by the application or service categories. The concept is to match the criticality of the application with the level of readiness service. Therefore, a safety-critical application should have a service of 99.9999% up-time, while applications where downtime has little effect on the organization can have 99% to 99.9% up-time. We now discard this theory and replace it with a simpler theory called the Black and White Availability Theory.

Median Economic Theory: Also known as sizing for the median. It is common practice for IT managers to size for the peak load and then add headroom. Most systems will have 20% to 30% headroom. In the Megaplex you size to the median level , which uses capacity on-demand and utility pricing. Customers will enjoy many benefits if they can size for the median. These benefits include fair pricing. The customer can buy the smallest system to meet its average workload without concern of under- or overcapacity. The overall cost of the system is dictated by actual usage. The more erratic the usage is, the fairer the cost to the organization. The customer also has the flexibility to balance costs between capital budgets and operating budgets. In our DARTS survey, 55% of IT executives said they are more likely to deploy internal cloud computing because it favors their capital budget. The same executives said it would also favor their operating budget.

	MORE	MORE	KSU
Megaplex			$11
IBMz Series	36%	$966	$18
Unix	35%	$901	$18
NonStop	26%	$582	$15
Linux	-15%	$233	$10

The chart shows the relative cost of a payment system on an IBM mainframe (zSeries), NonStop blades, Unix and Linux, and the Megaplex. In the IBM mainframe, NonStop blades, and Unix and Linux environments, 100% of the production is running on the platform. In the Megaplex example, 25% of the load is shifted to the NonStop environment, while 75% is left running on the low-cost Linux environment. In the above summary chart we show the cost differential of only 10%. This will be supported by the additional charts within.

MEGAPLEX MODELING

In the Road to the Megaplex, we stated that the most compelling reasons for modernizing NonStop applications and the NonStop environment are the sheer flexibility and cost-reward benefits. As applications decompose into services, each service can then be evaluated on an importance scale. The service also can be evaluated on availability needs, security needs, and data integrity needs without sacrificing the budget or technology strategy. Therefore, higher-importance services can operate on the higher-grade system services; the lower-importance services can operate on less costly service platforms. Our research and VirtualADVISOR show moving 25% of the workload to the NonStop from the lowest-cost Linux environment can increase the overall cost by only 10% while providing the highest level of service. Below we present the details of our theories and their related economic impact for organizations.

Now in its second decade of use, The Standish Group's VirtualADVISOR® System is a collection of proven wisdom-based management tools. The VirtualADVISOR® System uses highly advanced case-based reasoning technology. The Standish Group helps IT managers increase their understanding of their business and IT environment by providing case-based and enterprise-wide estimating solutions.

At the heart of the VirtualADISOR is the normalization process. This process converts all costs and work into thousands of Standish Units (KSUs). These KSUs can then be applied across different systems, applications, and workloads, allowing The Standish Group to have precise total cost of ownership (TCO) estimates with the greatest flexibility and accuracy. In this paper, we are estimating the total cost to operate a payment application with the exact same workload over various platforms as compared with the Megaplex. In these cases, the Megaplex is operating 25% of the workload in the NonStop mode and 75% in the Linux mode.

In Case 1: NSB v. Megaplex, we are measuring the NonStop BladeSystem in comparison to the Megaplex. The total cost per year to operate the NonStop BladeSystem is $2.3 million. The Megaplex drops that cost to $1.7 million, or a total savings of 26%.

In Case 2: Unix server v. Megaplex, we are measuring the cost of a Unix server operating in high-availability mode using a popular high-availability database with automatic failover and replication. The total cost per year to operate the Unix server is $2.6 million. The cost of Megaplex is the same as for case one, with $1.7 million or total savings of 35%. The Unix System is an increase in cost of 9% over the NonStop BladeSystem.

CASE 1: NSB v. MEGAPLEX

BASIC COST ($000)	HP-NSB	MEGAPLEX	DIFFERENCE
Hardware Cost	301	142	53%
Software Cost	110	33	70%
Manpower Cost	348	69	80%
Maintenance Cost	229	56	76%
Other Cost	217	66	70%
Total Basic Cost	1,205	366	70%
APPLICATION COST ($000)			
Basic Cost	1,205	366	70%
Software Infrastructure	300	138	54%
Database & Systems Admin	311	468	-50%
Application Maintenance	329	516	-57%
Other Cost	94	166	-77%
Total Operating Cost	2,239	1,654	26%
Cost of Downtime	31	34	-10%
Cost, Including Downtime	2,270	1,688	26%

CASE 2: UNIX SERVER v. MEGAPLEX

BASIC COST ($000)	UNIX	MEGAPLEX	DIFFERENCE
Hardware Cost	198	142	28%
Software Cost	77	33	57%
Manpower Cost	69	69	0%
Maintenance Cost	121	56	54%
Other Cost	92	66	28%
Total Basic Cost	557	366	34%
APPLICATION COST ($000)			
Basic Cost	557	366	34%
Software Infrastructure	343	138	60%
Database & Systems Admin	580	468	19%
Application Maintenance	746	516	31%
Other Cost	314	166	47%
Total Operating Cost	2,540	1,654	35%
Cost of Downtime	49	34	31%
Cost, Including Downtime	2,589	1,688	35%

KSU Costing Theory: Introduction

The VirtualADVISOR case database is always fresh and up to date; no case in the cost database is older than six months. Since the VirtualADVISOR System is a web-based, thin-client tool, new data is added and existing data is updated in the background. Therefore, each time a case is run, you are getting the most up-to-date results. Each TCO case in the database takes about 50 hours of work to complete, usually by several people. The TCO was developed by user input; all data is collected from users and no price data, performance, or any other input is derived from vendors. We do not use benchmarks or consultant opinions to calculate cost or risk. This makes the VirtualADVISOR a true unbiased source. All downtime data comes from primary research instruments and is collected from the actual users.

We continue our cases with a look at the IBM z10 and Linux compared with the Megaplex. In these cases the Megaplex is operating 25% of the workload in the NonStop mode and 75% in the Linux mode.

In Case 3: IBM z10 v. Megaplex, we are measuring the cost of an IBM z10 system operating in high-availability mode. The total cost per year to operate the IBM z10 server is $2.7 million. The IBM z10 cost is $1 million more than the Megaplex. The IBM z10 is an increase of 10% in cost over the NonStop BladeSystem.

In Case 4: Megaplex v. Linux, we are measuring the cost of a Linux system operating in Google-style high-availability mode, lots of cheap servers. The total cost per year to operate the Linux server is $1.5 million. The Linux cost is almost $200,000 less than the Megaplex, or 14% lower than the Megaplex including the cost of downtime.

CASE 3: IBM z10 v. MEGAPLEX

BASIC COST ($000)	IBM z10	MEGAPLEX	DIFFERENCE
Hardware Cost	229	142	38%
Software Cost	137	33	76%
Manpower Cost	461	69	85%
Maintenance Cost	94	56	40%
Other Cost	202	66	67%
Total Basic Cost	1,123	366	67%
APPLICATION COST ($000)			
Basic Cost	1,123	366	67%
Software Infrastructure	286	138	52%
Database & Systems Admin	395	468	-18%
Application Maintenance	661	516	22%
Other Cost	156	166	-6%
Total Operating Cost	2,621	1,654	37%
Cost of Downtime	33	34	-3%
Cost, Including Downtime	2,654	1,688	36%

CASE 4: MEGAPLEX v. LINUX

BASIC COST ($000)	MEGAPLEX	LINUX	DIFFERENCE
Hardware Cost	142	113	20%
Software Cost	33	14	58%
Manpower Cost	69	22	68%
Maintenance Cost	56	27	52%
Other Cost	66	40	39%
Total Basic Cost	366	216	41%
APPLICATION COST ($000)			
Basic Cost	366	216	41%
Software Infrastructure	138	80	42%
Database & Systems Admin	468	494	-6%
Application Maintenance	516	457	11%
Other Cost	166	159	4%
Total Operating Cost	1,654	1,406	15%
Cost of Downtime	34	49	-44%
Cost, Including Downtime	1,688	1,455	14%

MEGAPLEX MODELING

THE FUTURE OF NONSTOP DEMAND

Black and White Availability Theory: Cost of Downtime

In 1998, we published a paper titled Pound Foolishness, where we laid out the gradient of availability with the notion of rating the application and then matching it with the system. The scale in the paper had five levels for the application Standish Availability Dependency (SAD) grade. This ranged from a minor inconvenience to a very major problem. The second five-level scale was called the Standish Availability Competency (SAC) grade. This scale looked at the outage history of technology. The scale went from very major outage history to very minor outages. The idea was to rate the application on the SAD scale and match it to the SAC scale. A short time later we followed up by creating a single scale, ranging from task not critical to task safety-critical.

SAD/SAC and the task scale have served us well for more than a decade, but the time has come to retire the scale, forget about 9's, and replace all that with the Black and White Availability Theory. Here you checkerboard your application services and put them on either black or white. Application services that are on black are services that would cause the organization to go into the red if they were unavailable. Services that are on white will have a minor impact. You create two operating environments, one black and one white. The black environment will be at a high state of readiness. The white environment will take the Google approach to availability, which is lots of cheaper server blades running Linux and replacing them when they break.

We define the cost of downtime literally as the associated costs to an organization when an outage occurs. These figures include the cost to bring the system back to an operational state. Expenses also may include staff cost to make up for the loss in production during the outage, as well as direct and indirect lost revenue, including the cost to replace clients lost due to the outage. In the cost per minute of downtime chart, we see the average cost of a minute of downtime for online sales and orders to be $18,000. However, only 20% of the services of this application will cause the loss, whereas 80% of the services will cause little damage if unavailable. The cost to prevent downtime is also real. It will cost an organization three to four times more in TCO to have a system at the high state of readiness. With an integrated black and white system, the organization can run the black services on a black system and the white services on the white system. The Megaplex is the black and white system.

The table shows the average cost of a minute of downtime for some of the most popular applications. These costs are derived from customer data, survey data, and case data. The cost of a minute of downtime will vary by production load, peak versus off-peak, and other factors.

AVERAGE COST OF DOWNTIME

APPLICATION NAME	COST/MINUTE
Trading (securities)	$85,000
HLR	$35,000
Online Sales & Orders	$18,000
ERP	$16,000
CRM	$10,000
EFT	$7,500
E-Mail, Texting, IM	$5,000
ATM/POS	$4,000
Billing	$3,000
Social WebApps	$100

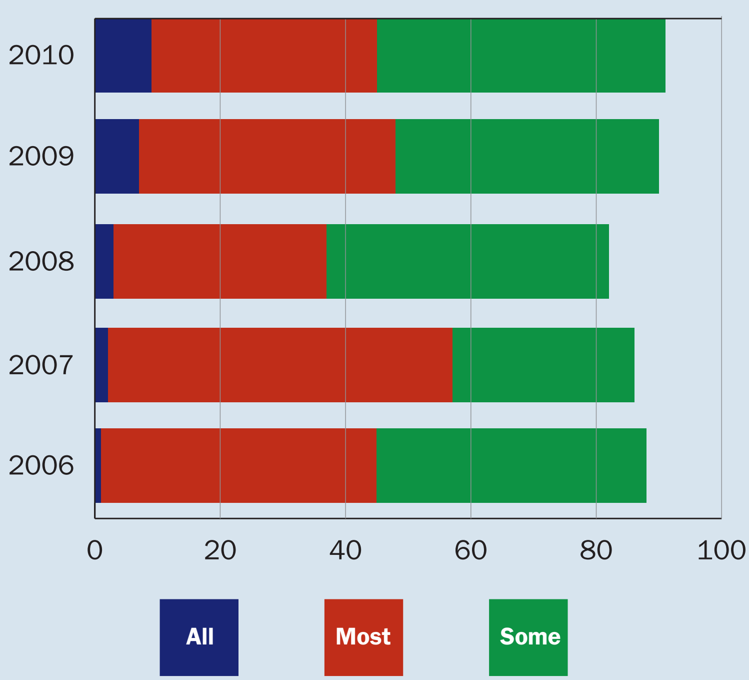
Black and White Availability Theory: Excess Capacity

Load and performance affect availability. Application performance is a function of many variables, including user demand, system capacity, and system tuning for the application(s). If the system is overloaded and business transactions cannot be entered or cannot be processed within acceptable or adequate response times, you'll have the equivalent of downtime in which associated costs need to be added. About 9% of organizations always have excess capacity to optimize availability, another 36% have excess capacity on most applications, and 46% have it on some applications. Almost 80% of organizations do some load testing, but for most of those it is done infrequently. The Excess Capacity chart shows the results when we asked DARTS respondents, "In general, do you maintain excess processing capacity to optimize for high availability?" With the Black and White Availability Theory there is no excess capacity—black capacity is available from white capacity. The real key is to turn white capacity into a high state of readiness.

More than 50% of IT executives favor cloud computing for server utilization. The reason for this is ready access to capacity on demand. A Black and White availability program must always provide for capacity ready resources. The Black and White Availability Theory goes beyond other programs where capacity is on demand, but requires manual intervention and approval. With black and white availability there is no intervention or manual process. If the application requires more capacity, it uses it. In addition, the program should work in hour increments. There is a base level utilization amount (KSUs) and anything over the base-level utilization is reported and charged by the hour/KSU, not on the less precise and larger granular monthly, or worse, quarterly basis. Occasional bursts of usage and the need for acceleration capacity are supported by the Black and White Availability Theory. The Megaplex provides for excess capacity without human intervention.

THE STANDISH GROUP

EXCESS CAPACITY



	2006	2007	2008	2009	2010
Some Apps	43%	29%	45%	42%	46%
Most Apps	44%	55%	34%	41%	36%
All Apps	1%	2%	3%	7%	9%

The chart shows the results when we asked DARTS respondents, "In general, do you maintain excess processing capacity to optimize for high availability?" In 2010, 91% of organizations had at least some applications with excess capacity to accommodate availability.

MEGAPLEX MODELING

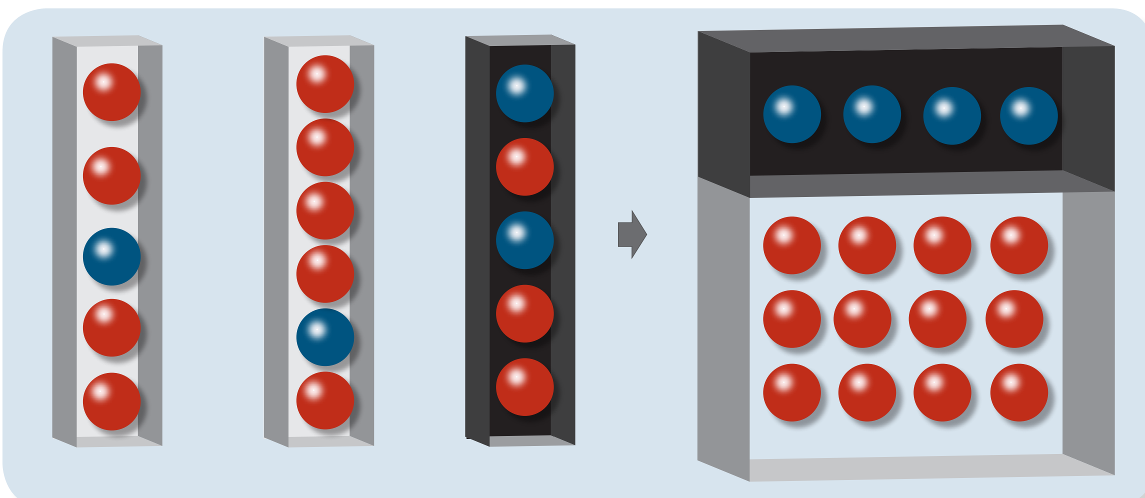
Black and White Availability Theory: SOA

Seventy percent of IT executives believe a service-oriented architecture (SOA) will provide them with at least some relief from executive demands. These same 70% are using open-source objects to augment existing applications or services. Almost 90% of IT executives believe that an SOA can speed up development. Almost 40% of organizations have an active program to break up stovepipe applications into services. The Megaplex offers these organizations the ability to move critical parts of the application onto a system with a high state of readiness, yet only pay for actual usage. As the move to the Megaplex draws in other resources with less demanding requirements onto the commodity side, the result is a very cost- effective solution.

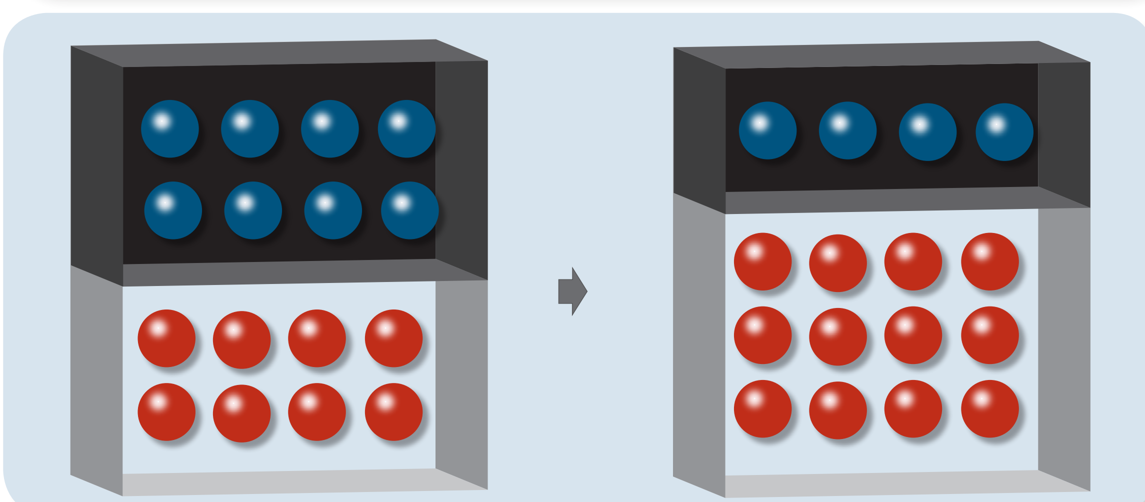
In Figure A below we see red and blue services. The left-hand side shows how these services are currently deployed. The white vertical boxes indicate common Unix/Linux systems and the black vertical box indicates a high-availability system. The right-hand box shows how the blue services fit within the high readiness of the Megaplex and the red in the commodity part of the Megaplex.

In Figure B we see the blue services taking more resources and squeezing out the less valuable red services. The organization will only pay for the time these blue services are in the black zone. When the workload goes back to average level the system will return the resources to the red services.

A



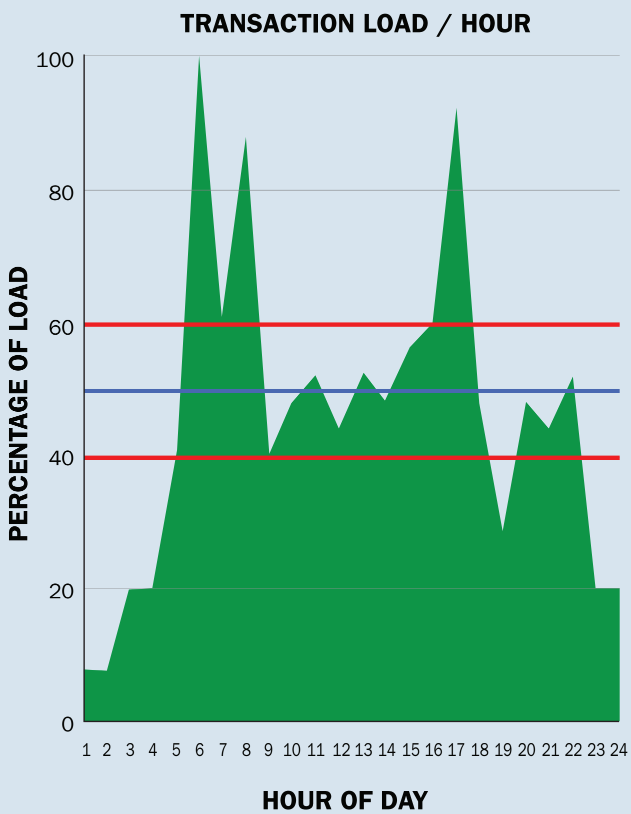
B



Median Economic Theory

Many years ago one of the founders of The Standish Group owned an outsourcing company. The company provided manual IT services such as programming and data entry. Much of the work the company received was from other organizations' overload that the current staff could not handle. In order for the outsourcing company to expand sales, the company came up with what it termed the Median Economic Theory. The theory states that in order to optimize cost, the organization should always staff to the median level and pay the premium for the overload.

In any cycle there will always be peaks and valleys. The real issue around the median is the difference between how low are the lows and how high are the highs. If the median level is set too low, the price of premiums could cause the overall cost to go higher. If the median is set too high, then the cost of waste would be higher. The real trick is to get the right median level.



The chart shows the typical daily cycle of a transaction processing application. The blue denotes the median and the red lines are upper and lower median.

The Standish Group has coined the phrase "NonStopDemand (NSD)" to describe a potentially new pricing option. NSD incorporates the concept of "Fair Pricing" by charging the users for their actual use and encouraging higher use by offering workload savings. This new capability would reduce the barrier to move critical application services onto a readiness platform and make the Megaplex even more competitive. The below chart shows the cost of the same workload as in cases 1 to 4, but charging a base fee and monthly usage charges. The cost for the Megaplex with NSPricing is $10.5 thousand per KSU, almost the same cost as the 100% commodity platform running Linux.

CASE 5: MEGAPLEX v. NSDEMAND - MEDIAN LOAD

BASIC COST ($000)	MEGAPLEX	NSDEMAND	DIFFERENCE
Hardware Cost	142	92	35%
Software Cost	33	17	48%
Manpower Cost	69	69	0%
Maintenance Cost	56	56	0%
Other Cost	66	66	0%
Total Basic Cost	366	300	18%
APPLICATION COST ($000)			
Basic Cost	366	300	18%
Software Infrastructure	138	90	35%
Database & Systems Admin	468	468	0%
Application Maintenance	516	516	0%
Other Cost	166	166	0%
Total Operating Cost	1,654	1,540	7%
Cost of Downtime	34	34	0%
Cost, Including Downtime	1,688	1,574	7%

MEGAPLEX MODELING

In Summary

In most every application there are parts that could use a highly available and secure platform. This platform must ensure data integrity and be highly scalable. The platform of the future must be capacity-ready without requiring manual intervention. The platform must provide for usage billing, but have the flexibility to alter the billing as the median workload increases or decreases. As shown in the table below, the Megaplex with NSDemand billing can come within 8% of a commodity platform and even enjoy lower hardware cost.

Modeling the Megaplex is an interesting exercise. The models in this report are examples of the types of cost savings an organization could enjoy. Modernizing the applications as laid out in the Road to the Megaplex can bring enormous benefit in terms of both cost and service levels. Certainly Linux is the operating system of both the present and the future. On the other hand, the world still needs a system that offers superior availability, scalability, and security.

CASE 6: NSDEMAND v. LINUX

BASIC COST ($000)	NSDEMAND	LINUX	DIFFERENCE
Hardware Cost	92	113	-23%
Software Cost	17	14	18%
Manpower Cost	69	22	68%
Maintenance Cost	56	27	52%
Other Cost	66	40	39%
Total Basic Cost	300	216	28%
APPLICATION COST ($000)			
Basic Cost	300	216	28%
Software Infrastructure	90	80	11%
Database & Systems Admin	468	494	-6%
Application Maintenance	516	457	11%
Other Cost	166	159	4%
Total Operating Cost	1,540	1,406	9%
Cost of Downtime	34	49	-44%
Cost, Including Downtime	1,574	1,455	8%

Please note: *The Megaplex is not a real product, but a product we envision HP will be offering in the near future. In addition, utility billing, as of the writing of this paper, is not offered by HP. If you would like HP to offer this billing option, please contact your HP representative. The Standish Group is available to model your workload and provide total cost of ownership and return on investment services for the Megaplex and other platforms. Megaplex Modeling: the Future of NonStop Demand is based on a number of primary research instruments, including Initial CENTS (Comparative Economic Normalization Technology Study), Monthly CENTS on downtime, DARTS (Demand Assessment Requirements Tracking Studies) interviews, and workshops. All research participants must satisfy a qualification process and join our Standish User Research Forum (SURF). It should be noted that none of the predictions of future products come from inside or confidential information. These predictions are simply the imagination of The Standish Group and and in some cases, our recommendations to HP. All data and information in this report should be considered Standish opinion, and the reader bears all risk in the use of this opinion.*

The Standish Group International, Inc.
60 State Street, Suite 700
Boston, MA 02109
P: (508) 760-3600
www.standishgroup.com