Storage Validation at Go Daddy

Best practices from the
world's #1 web hosting provider

March 2013

Table of Contents

Abstract

Faced with escalating data volumes and storage costs, and the promise of Infrastructure-as-a-Service, IT organizations must innovate. To enable its innovation, Go Daddy, the world's leading web hosting services provider has established best practices for storage technology validation. Supported by testing and analytics solutions from Swiftest, the industry leader in storage validation, the new process empowers Go Daddy with the insight they need to control storage costs and optimize service delivery.

Readership

This paper is for data storage professionals and IT management at large enterprises and service providers, or at any company with a large storage infrastructure. Engineers and architects will be interested in the validation approaches described and the examples of findings shared by the Go Daddy team. Decision makers will consider the robust business case for implementing such a process.

Introduction

These are good times for storage professionals. A flurry of technologies, from SSDs to object storage, caching/tiering, or dedup/compression, promise faster, cheaper and better storage solutions. Storage-as-a-service models offer a new blueprint for flexible, optimized storage operations. The opportunities for cost savings from these innovations are considerable, and tantalizing to many, in the face of escalating data volumes and storage costs. Yet many IT organizations struggle to address these opportunities.

Meet the Go Daddy storage team. They manage a 27PB storage infrastructure with 99.999% uptime, in a storage-as-a-service model. Their technology roadmap will reduce their unit storage costs by over 65% over the next two years. How is this possible? It all starts with a robust engineering process. This paper describes Go Daddy's storage validation process, which allows them to rapidly evaluate new technologies and innovate with confidence. You'll learn about best practices through Go Daddy's experience; what they did, learned, and gained.

Go Daddy's Storage Challenge

Business Highlights

Go Daddy is the world's largest hosting service provider with a mission "to enable individuals and businesses to establish, maintain and evolve an online presence." Here are a few highlights that illustrate Go Daddy's market leadership and operational scale.*

- Over 11 million customers worldwide
- Over five million hosting accounts — world's largest provider
- Over 54 million domain names under management and 50%+ market share
- Over 3,000 employees and 600 in-house developers
- 24 x 7 x 365 customer support—all US-based
-

* For more information, visit www.godaddy.com.

> "Load DynamiX puts us in the driver's seat when it comes to our storage roadmap and cost structure."
>
> **Justin Richardson**
> **Senior storage engineer**
> **Go Daddy**

Storage at Go Daddy

Go Daddy's storage teams manage a large (27PB) and fast growing (30% annual growth) infrastructure, distributed across five data centers globally. With three engineers and five administrators for this large an infrastructure, productivity is well above industry benchmarks.

Key highlights:

- Storage-as-a-Service delivery model
- 99.999% uptime SLA
- 95% unstructured data (NAS) vs. 5% SAN
- High I/O and low latency storage workloads

With millions of dollars spent every year in storage investments, unit cost reduction is a core objective. So is performance improvement, a key driver of customer experience. The storage team is consistently challenged with finding faster, better and cheaper storage solutions. The engineers are pressured to innovate, evaluate the latest and greatest technologies, and bring ideas to fruition rapidly. Recent engineering validations include SSDs, caching and tiering, deduplication and compression, as well as open compute on commodity hardware solutions.

The rest of this paper addresses how Go Daddy assesses and validates storage technologies, what they test today, the insights they've gathered, and the impact of these insights.

Go Daddy Validation Process Evolution

Go Daddy's validation process evolved through several stages. [Figure 1]
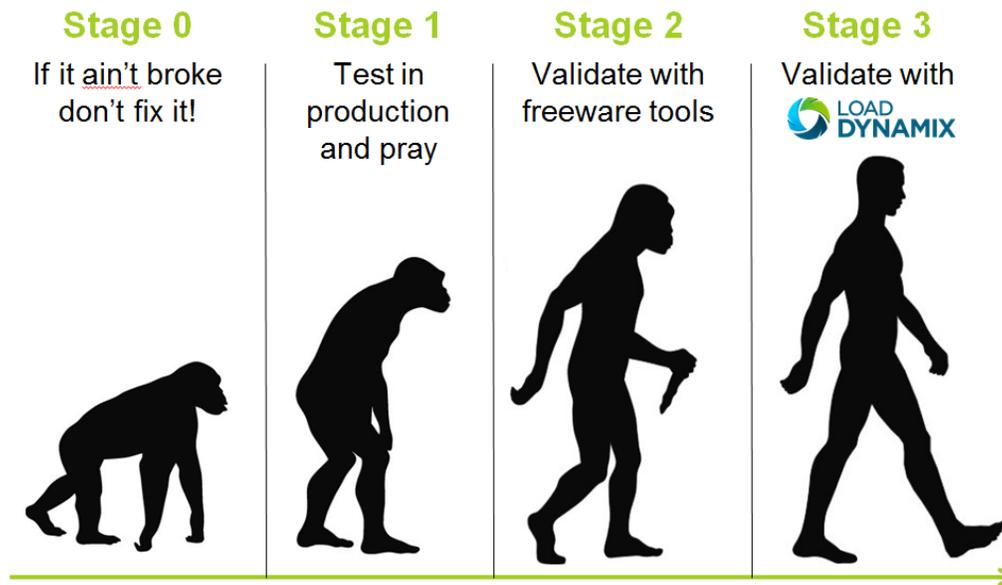

Figure 1: Evolutionary stages of storage validation at Go Daddy

Stage 0: If it ain't broke, don't fix it!

Go Daddy began their journey with a fairly common engineering outlook: risk averse and reactive. User experience being paramount, they addressed the unknowns in their storage environment by over-engineering and over-provisioning storage solutions, and paying high prices for premium branded products. This approach worked for some time, but left the team with limited insight into their solutions, unable to address change, and facing rapidly escalating costs. Unfortunately, many IT organizations remain at this stage today.

Stage 1: Test in Production

Pressures to innovate led the team to introduce new solutions. The next stage in Go Daddy's validation evolution was indeed to test, but test in production. New storage solutions were introduced with system design based on vendor spec—all by the book, with a controlled environment and slow customer ramp-up.

Careful, controlled increase in customer load is a common strategy, but it back-fired, as it delayed discovery of load-related failures for many months, and exposed customers to these failures. Without the ability to rapidly assess new solutions, engineering innovation was stunted. This approach cost Go Daddy heavily in time and resources, resulting in a commitment to fully test prior to production (and never on live customers!).

Stage 2: Validate with Freeware

Next stop on the validation journey was to test pre-production using freeware tools. There are many commonly used tools out there for offline validation. Go Daddy tried them all, but found that this approach did not significantly improve results.

Two key issues remained:

1. Scale: the team struggled to emulate the high-density workloads of its production environment.
2. Realism: tests created with these tools couldn't emulate metadata and many other file system calls.

This meant that Go Daddy could not evaluate the very large-scale failure scenarios it wanted to eliminate. And finally, the tools required costly, hard-to-manage high-end servers and were altogether cumbersome to use.

Stage 3: Validate with Load DynamiX

Through the trials above, Go Daddy learned just what's required for fast, continuous improvement of their storage infrastructure:

- Highly realistic emulation of their high-density storage workloads in a simple lab environment.

- Ability to find the failure limits of each solution.

- Precise and consistent measurements and findings.

- Results in days, not months.

These basic necessities and more were achieved with adoption of the Load DynamiX storage validation solution, a single solution for designing, running, and analyzing high performance tests. Go Daddy created a simple storage validation lab *[Figure 3]* and perfected a validation process that is now at the core of their storage engineering process *[Figure 2]*.

> "If you can't validate technology before it's deployed into production, then you are flying blind."
>
> **Julia Palmer**
> **Performance engineering manager**
> **Go Daddy**

Go Daddy Validation Process

Today, Go Daddy follows these steps [Figure 2] for validating new technologies in their Load Dynamics lab before release into production.

1. **Investigate:** Investigate new technologies, new features, or potential configuration changes. Scope out new systems, software, and vendors. Assess costs.
2. **Procure and install:** Get a full-scale POC system into the lab working with vendors or integrators.
3. **Apply workloads:** Use the Load Dynamics appliance to apply simulated production workload. Test until you get to pass/fail.
4. **Analyze, report, modify:** Use results to adjust configurations repeatedly based on results. Do you scale back on density? Increase performance? Go back to step 3.
5. **Release to production:** By this point, you're finely tuned and you can release your pilot to real customers. With confidence.

Go Daddy validation process

Today, Go Daddy follows these steps *[Figure 2]* for validating new technologies in their Load DynamiX lab before release into production.

1. **Investigate:** Investigate new technologies, new features, or potential configuration changes. Scope out new systems, software, and vendors. Assess costs.

2. **Procure and install:** Get a full-scale POC system into the lab working with vendors or integrators.

3. **Apply workloads:** Use the Load DynamiX appliance to apply simulated production workload. Test until you get to pass/fail.

4. **Analyze, report, modify:** Use results to adjust configurations repeatedly based on results. Do you scale back on density? Increase performance? Go back to step 3.

5. **Release to production:** By this point, you're finely tuned and you can release your pilot to real customers. With confidence.
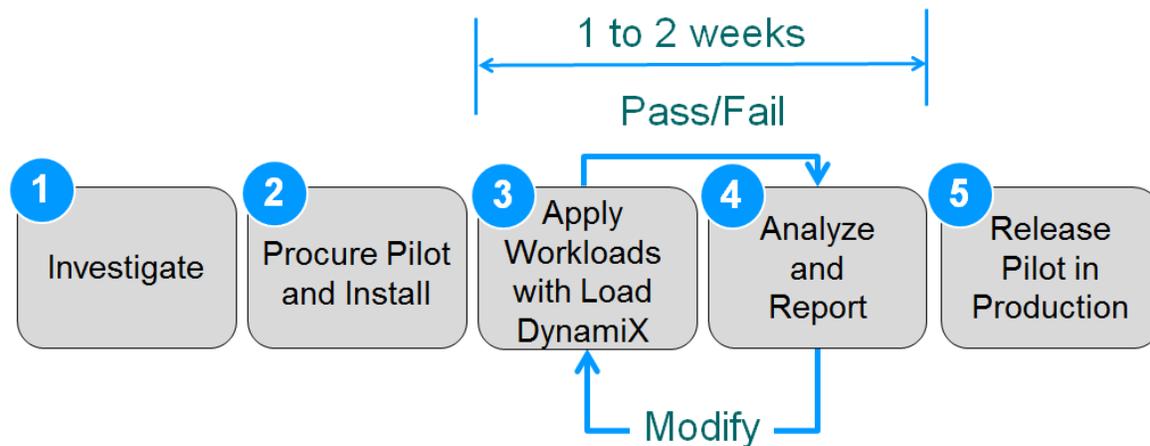

Figure 2: Go Daddy 5-step validation process

It now takes Go Daddy one to two weeks to cycle through Steps 3 and 4 and determine solution behavior at scale with confidence. It previously had taken Go Daddy up to 12 months to do so. The result is a fast-iterating engineering process that rapidly generates insight upon insight.

Use Cases

Go Daddy uses its storage validation process on a daily basis, for a variety of testing scenarios such as:

- **New feature evaluation:** Deduplication, compression.
- **Protocol evaluation:** Comparisons between NFS versions or, for example, CIFS/SMB vs. iSCSI.
- **Product bake-offs:** Comparisons among potential products and vendors.
- **New architecture/technology evaluation:** SSD, caching, tiering, open compute on commodity hardware.
- **Density testing:** Consolidation challenges? Great way to figure it out how many customers could be deployed on one box.
- **Routine change management:** Simple things. Are OS and firmware upgrades working?

Go Daddy Lab Setup

The Go Daddy lab layout is simple by design to ensure a controllable environment.

The Load DynamiX appliance is at the core of the set-up, used as a load generator and connected to a 10GE switch *[Figure 3]*. The current production solution is used for validating OS, firmware and for configuration updates. Go Daddy can quickly swap in and out new candidate solutions (shown as A and B) for POCs and test them against an emulation of Go Daddy's production workload. The Load DynamiX appliance is controlled from the desktop so Go Daddy can push loads and change settings remotely.
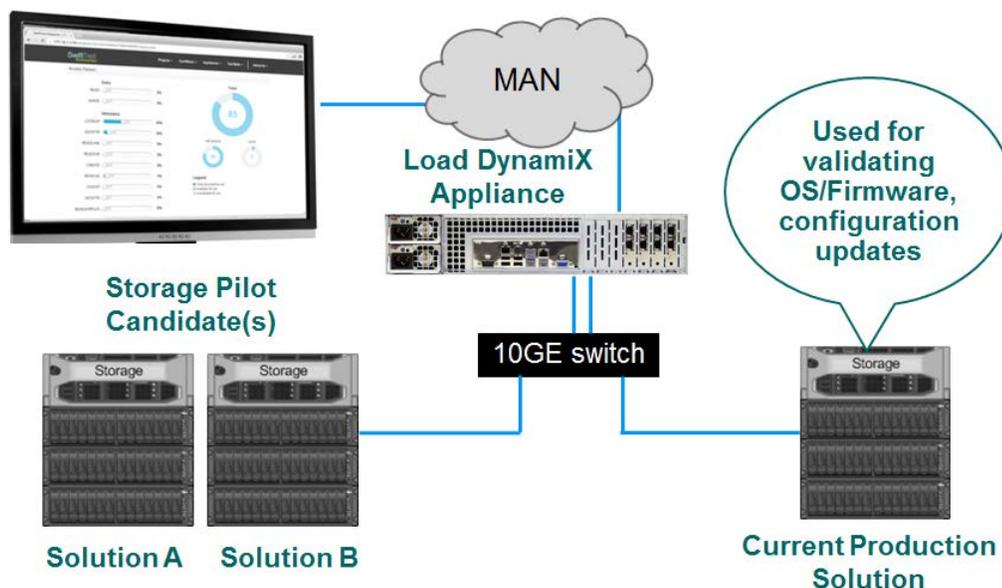

Figure 3: Go Daddy's Load DynamiX lab layout

A strong, capable engineering team is key to any engineering process. The Go Daddy team includes 2 NAS storage engineers and 1 SAN storage engineer. They are eager to investigate leading edge storage technologies and keep their focus on the complete picture (cost, performance, availability). Storage engineering expertise is important: while the Load DynamiX solution is designed to be easy to use and intuitive, users should be knowledgeable about storage technologies and comfortable with storage protocols and file system concepts.

Dialing in Your Workload

To implement its storage validation process, Go Daddy needed first to emulate its production workload in the Load DynamiX application. For successful storage validation, it's essential that a company knows its workload, and test with a realistic emulation of it. Realism is the key!

How Go Daddy analyzes workload

Go Daddy used several different tools to obtain workload characteristics, including vendor-provided performance analysis utilities such as netstat and nfsstat. Such tools are available from several vendors and allow users to characterize their workloads with key I/O statistics such as:

- Bandwidth utilized
- Latency
- Read/write percentages
- Meta-data call percentages

Again, it is most critical that engineers characterize their workloads effectively. Small random, or large sequential? What percentage metadata, etc.?

In a comparative analyses *[Figure 4]*, see the characterization of Go Daddy's actual NFS production workload (left) and the Load DynamiX emulated workload (right). They are strikingly similar, between a 90% and a 100% match.

	Go Daddy Production	Load DynamiX Simulation
Total NFS ops	~65K	~66K
Avg. Latency	1.5 ms	1.4 ms
▪ Read	10.0 ms	11.5 ms
▪ Write	0.5 ms	0.6 ms
▪ Metadata Ops	0.5 ms	0.4 ms
Op-Mix		
▪ Reads	5%	5%
▪ Writes	1%	1%
▪ Metadata Ops	94%	94%
▪ Getattr	62%	63%
▪ Lookup	11%	13%
▪ Access	17%	14%
Avg. CPU Utilization	81%	80%
Max. Disk Utilization	55%	54%

Figure 4: Comparing production with simulation

See that IOPS were nearly matching, at around 65,000. Latency was also similar. The mix of NFS calls was key—Go Daddy wanted to align reads and writes as closely as possible.

In addition, the team saw that Metadata OPS represented a large percentage of overall workload so it was important to characterize this behavior. Note that those calls were also reproduced very faithfully by the Load DynamiX emulation.

Finally in the comparison, see that CPU and disk utilization are also close which indicates that Go Daddy is pushing lab hardware as hard as they're pushing production hardware.

Go Daddy adjusted the NFS calls using the Load DynamiX application until they had the right mix. Once the mix was dialed in, they increased the load until they matched full production.

Validation Examples

Go Daddy has conducted many different types of validation. Here are three examples in detail.

Hardware configuration

Go Daddy measured the feasibility of using lower cost disk solutions *[Figure 5]*. The current configuration (baseline) was running on 15K RPM drives. Two new configurations were compared; one using 7.2k RPM drives and the other a Tiered HDD with SSD.

With a test run at 35k IOPS, the new configurations showed latency at 20ms and 15ms respectively compared with 9ms baseline. And when load was dialed up to 70k IOPS and baseline latency was 10ms, the new configurations showed unacceptable latency of 30ms and 18ms. These hardware options proved not a good choice for Go Daddy, despite cost savings.

Configuration	Current Production 15k RPM Drives	New config. A 7.2k RPM Drives	New config. B Tiered HDD with SSD
Cost	baseline	45% decrease	15% decrease
Latency @ 35k IOPS	9ms	20ms	15ms
Latency @ 70k IOPS	10ms	30ms	18ms

Figure 5: Validating new hardware configurations

Caching, Tiering, and SSD Technologies

Go Daddy performed Pass/Fail validations for four different storage solutions. *[Figure 6]*

Solution A: Pass. This was a caching solution internal to the box. Go Daddy found that it offloaded 45% of their reads from the drives and they were also able to test cache "warm-up" time. Both are extremely useful.

Solution B: Fail. Software was found to be incapable of pushing high IO load on full SSD arrays to capacity. In this test, findings were quickly substantiated using alternate software, which aided in working with the vendor.

	Pass/Fail	Findings
Solution A	PASS	▪ Cache offloaded ~45% of READS from disks ▪ Able to test cache "warm-up" time
Solution B	FAIL	Software issue: software not capable of pushing the high IO load on full SSD arrays
Solution C	FAIL	Latency issues when SSDs were trying to flush writes to slow disk
Solution D	PASS	Full SSD array was able to push SSDs to max throughput.

Figure 6: Validating caching, tiering, and SSD technologies

Solution C: Fail. Latency issues were found when the SSDs tried to flush the writes to the disk. Go Daddy discovered that the workload for this solution was slightly random, so writes to disk were slow.

Solution D: Pass. A full SSD array was tested and successfully pushed to maximum throughput. This test not only validated a super fast array, it showed Go Daddy that they now had a load-test solution that can push enough I/O to storage to truly stress any system.

An unforeseen benefit of the flexible, fast-to-results testing that the Load DynamiX lab provides is increased collaboration with solution vendors. Vendors can be invited to watch testing and participate in reproduction of issues with measurements they can use to improve their products.

Open source software on commodity hardware

This test compared two solutions based on commodity hardware and open source storage software to the Go Daddy production reference. *[Figure 7]*

Go Daddy tested various drive speeds and weighed the costs for each of the two solutions running on those hardware options.

Note that both commodity/open source solutions offer significant performance improvements and cost reduction opportunities.

Solution 2, in particular, is much faster and much cheaper than the current production standard. So. Cheaper? Yes. Faster? Yes. But, are they better?

These solutions lack some of the features of the leading commercial solutions, as well as the technical support. Their stability is, as of yet, unknown. So, they may not be better.

However, Go Daddy's validation process showed that solution 2 is a great candidate for a pilot program—it ticks the boxes for faster and cheaper, and is worth exploring for overall advantage.

> "We can now assess the hottest storage technologies like SSDs, caching, tiering, and dedup, against our full production requirements, faster and more accurately than ever before."
>
> **Justin Richardson**
> **Senior storage engineer**
> **Go Daddy**

	Faster			Cheaper	Better
	15k RPM Drives	7.2k RPM Drives	Tiered HDD with SSD		
Vendor-A (production reference)	9ms	20ms	15ms	baseline	+ Features + Support + Mature/stable
Open Compute solution 1	8ms	18ms	N/A	23% reduction	− Support − Features − Stability unknown
Open Compute solution 2	6ms	13ms	15ms	56% reduction	− Support − Features − Stability unknown

Figure 7: Validating commodity hardware and open source software solutions

Conclusion

The implementation of Load DynamiX storage validation for Go Daddy has made an extraordinary impact on their storage engineering process.

Armed with new insights, the team has laid out a technology roadmap that will reduce unit storage costs by over 65% over the next two years *[Figure 8]*, resulting in massive overall savings. Every technology transition along the way will be validated using their storage validation process with Load DynamiX, giving the team full confidence that the each change will   work in production.
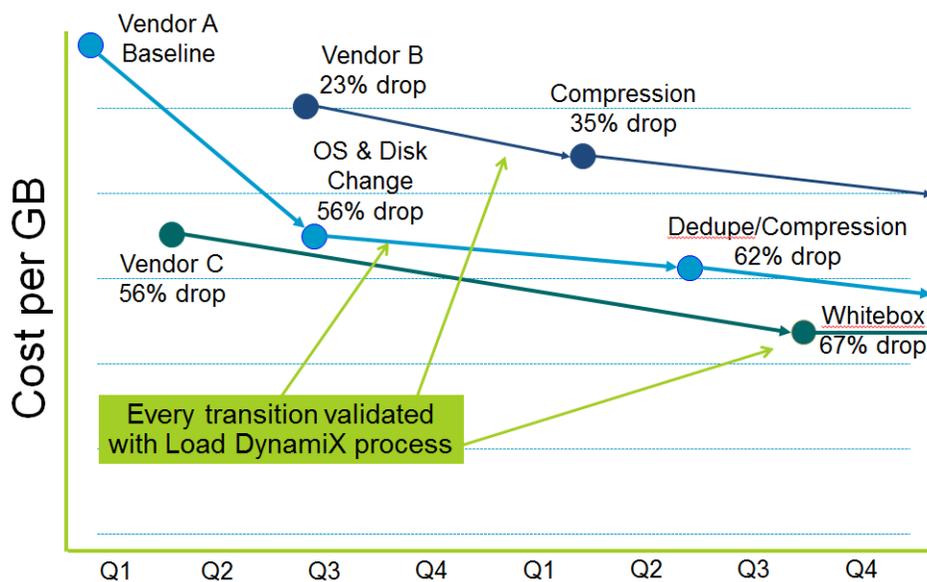

Figure 8: Go Daddy cost reduction two-year forecast

Go Daddy's testing best practices boil down to:

- A rigorous understanding of their workload
- A drive to investigate the leading edge while keeping a balanced perspective of costs, performance and availability
- Systematic evaluation of new solutions through realistic and comprehensive validation, made possible by Load DynamiX.

Proven benefits for Go Daddy:

- **Price/performance:** Choice of the most cost effective solutions for Go Daddy's particular workload
- **Performance in production:** Confidence in a positive user experience
- **Availability:** Confidence in five 9s under load
- **Freedom to innovate:** Insights that power a robust engineering process
- **Time to validate:** Weeks to validate in the lab instead of 12 months in production

In this paper, we have shown how pre-production validation in their Load DynamiX lab has enabled Go Daddy to confidently deliver on core promises to the business.

To listen to a webcast covering the content in this paper, visit http://www.LoadDynamiX.com and go to resources section under videos.

> "Load DynamiX gives us the confidence of knowing exactly how our infrastructure will behave in the real world, in a matter of days instead of months. It helps us determine an optimal balance of better, faster, and cheaper storage system technologies that ultimately benefit Go Daddy customers."
>
> **Julia Palmer**
> **Performance engineering manager**
> **Go Daddy**

Load DynamiX Product Information

The Load DynamiX storage infrastructure performance validation product suite is comprised of workload modeling software that helps IT managers capture data and statistics on their installed applications to create realistic workload models and load generation appliances that deliver extreme workloads to 1G, 10G, or Fibre Channel storage targets.

The Load DynamiX product suite delivers:

- **A comprehensive storage validation solution:** Unify storage validation processes with a single test product for File, Block, and Object / Cloud storage protocols such as NFS, SMB, iSCSI, FC, CDMI, and OpenStack Swift.

- **Emulation of real-world traffic:** Get the most accurate and efficient workload simulation with the industry's deepest storage protocol emulations.

- **Ease of use and fast time to value:** Start validating storage

  systems quickly with an intuitive set-up and a library of prebuilt tests.

Figure 9: The Load DynamiX product solution comprises both software and hardware components.

About Go Daddy

Go Daddy is the world's largest domain name provider, Web hosting provider and new SSL provider, focused on helping small businesses grow larger. Go Daddy provides dozens of cloud-based services and is the largest worldwide mass-market hosting provider by annual revenue according to 451 Research (Mass-Market Hosting Report-Fall 2012) and is the #1 provider of net-new SSL certificates for 2012, according to the Netcraft, LTD Secure Server Survey.

About Load DynamiX

As the leader in infrastructure performance optimization, Load DynamiX empowers IT professionals with the insight needed to make intelligent decisions regarding networked storage. By accurately characterizing and emulating real-world application behavior, Load DynamiX optimizes the overall performance, availability, and cost of storage infrastructure. The combination of advanced workload modeling and performance validation software with extreme load-generating appliances give IT professionals the ability to cost-effectively stress today's most complex physical, virtual and cloud infrastructure to its limits.

Visit www.LoadDynamiX.com for more information.