

# Storage validation at GoDaddy

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

Julia Palmer

**Storage and Backup Manager**

Justin Richardson

**Senior Storage Engineer**



# Agenda and Benefits

- ❑ Agenda
  - ❑ **Faster, better, cheaper** storage challenges
  - ❑ Evolution of our validation process
  - ❑ Insights and benefits
- ❑ **Why will you learn?**
  - ❑ **How do we test new storage technology and achieve:**
    - ❑ Storage cost reduction
    - ❑ Confidence in user experience (performance and 99.999 storage availability)
    - ❑ Freedom to innovate!

# GoDaddy key facts

- ❑ **GoDaddy enables individuals and businesses to establish, maintain and evolve an online presence**
  - ❑ Nearly \$1.3 billion (B) in sales
  - ❑ 12 million customers worldwide,
  - ❑ World's largest Web Hosting Provider, with over 5.6 million hosting accounts
  - ❑ 99.9% uptime guarantee for Web hosting
- ❑ **Key storage infrastructure highlights**
  - ❑ **31+ Pbytes**, 30% annual growth
  - ❑ 95% unstructured (NAS), 5% structured (SAN)
  - ❑ Distributed over 5 data centers on 3 continents
  - ❑ 5 storage administrators (more than 6PB per admin)
  - ❑ 3 storage engineers

# Storage Technology infrastructure challenges

- ❑ **99.999% uptime for storage services.**
  - ❑ Whatever we design must work in production.
  - ❑ Minimal downtime, planned or unplanned.
- ❑ **Cost reduction**
  - ❑ **Provide Storage customers with better performance while reducing our costs.**
- ❑ **Innovate!**
  - ❑ Find FASTER and BETTER and CHEAPER technologies.
  - ❑ Leverage new features, protocols, hardware, tools.
- ❑ **Go fast!**
  - ❑ Speed up process from idea to production release
  - ❑ Ensure efficient engineering processes for POC and Pilots
  - ❑ Automate validation and QA

# Our topic today

**Faster, Better and  
Cheaper** storage  
services?  
**How?**



Storage  
Engineer

# Storage technology innovations!

Cool storage  
technology?

SSDs

Deduplication

Virtualization

NFS4.1,  
SMB 3.0

Caching/Tiering

Unified  
Storage

Compression

Object  
Storage

Storage  
Engineer

Commodity  
Hardware



# Storage technology innovations

Would it  
**work in  
production?**

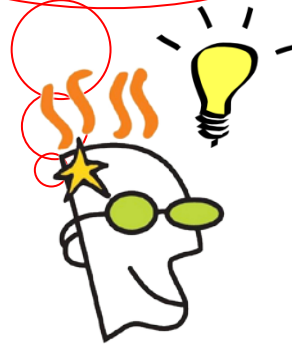
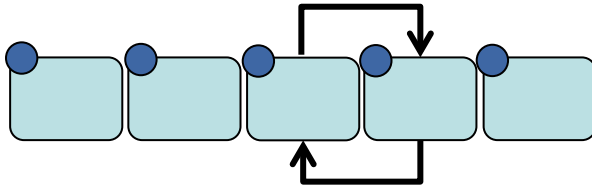
How can I  
find out?



Storage  
Engineer

# Storage technology innovations requirement #1

Storage validation process with a  
validation solution!!!



Storage  
Engineer

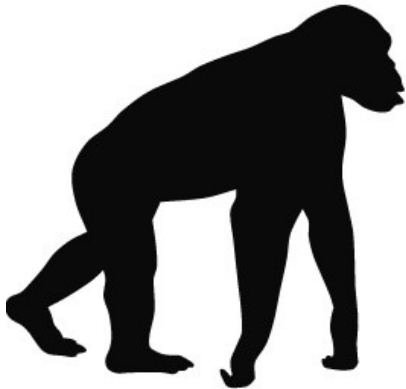


# Stage 0: Storage validation process evolution

## Stage 0

*“If it ain’t broke don’t fix it!”*

- Engineering process is risk averse. Difficult to investigate new solutions – primarily reactive to problems.
- High costs!
  - Over-spec solution in order to accommodate the unknown
  - High priced, premium branded product
- Cannot keep up with changes in demand
- Limited insight hindering ability to innovate



# Stage 1: Storage validation process evolution

## Stage I

*“Test in Production... and Pray”*

- ❑ Introduction of new storage solutions to production in controlled environment with slow customer ramp up
- ❑ Design based on vendor specs
- ❑ Load-related failures may take 12 months to happen
  - ❑ Highly visible failures (potentially thousands of customers disrupted)
  - ❑ 12 months wasted
- ❑ CIO directive: ***“No more testing on live customers!!!”***



# Stage 2: Storage validation process evolution

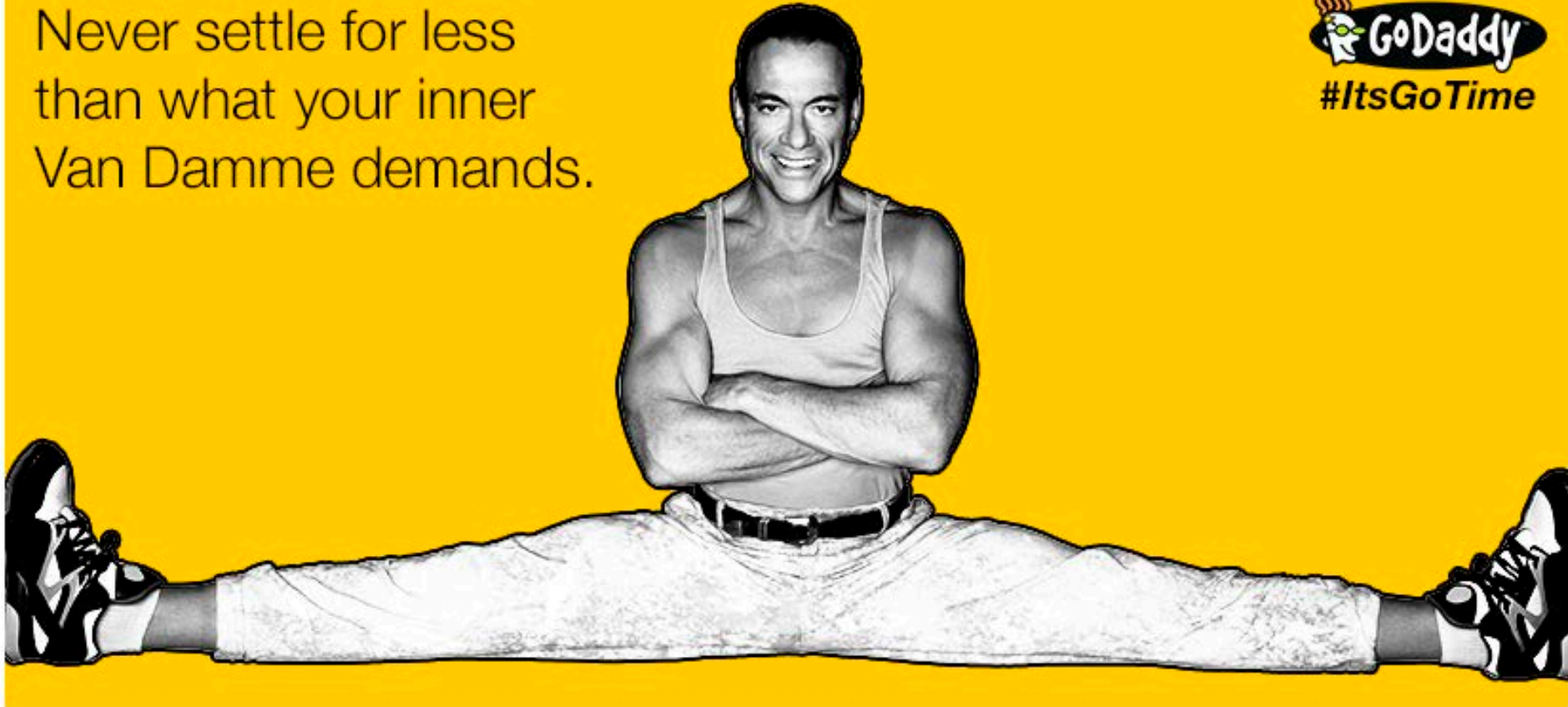
## Stage 2

*“Validation process with freeware tools”*

- ❑ Offline validation of candidate solutions with mix of tools: IOMeter, IOZone, Dbench, Fstress, etc...
- ❑ Could not scale up to our high density production loads – so could not address the **large scale failure issue**
- ❑ **Lack of realism:** inability to emulate meta-data and other file system calls.
- ❑ Required lots of high end servers
- ❑ Cumbersome to use



Never settle for less  
than what your inner  
Van Damme demands.

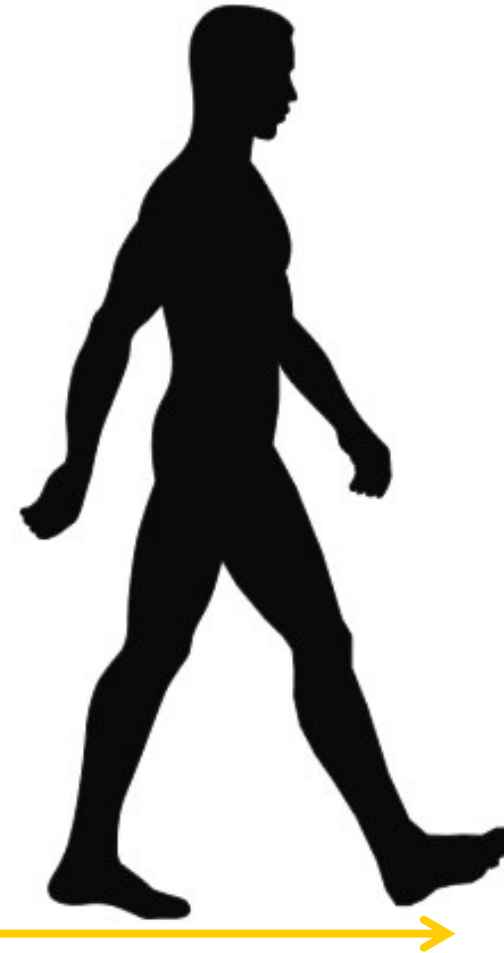


# Stage 3: Storage validation process evolution

## Stage 3

### *Validation process with **SwiftTest Appliance***

- ❑ Rapid evaluation of storage configurations
  - ❑ Find out in days not months!
  - ❑ Failure is not an option – it is a requirement.
  - ❑ Precise and measurable findings!
- ❑ Test solution
  - ❑ Realistic emulation of our production workloads (~90%+ fidelity, metadata calls, etc.)
  - ❑ Validates against full scale operation loads
  - ❑ Single solution for designing, running and analyzing tests
- ❑ Flexible, lightweight lab - sandbox for designing, running and analyzing different HW, SW and firmware options

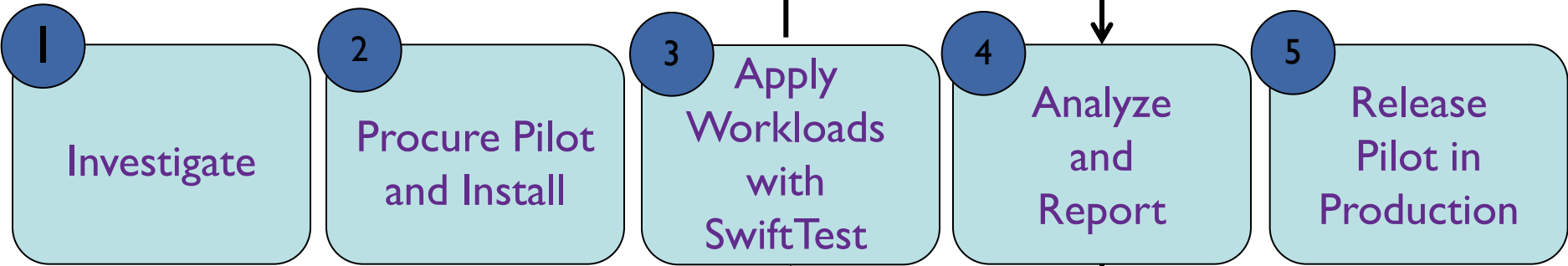


# GoDaddy storage validation process

1 to 2 weeks

Pass/Fail

Modify



- ❑ Technology
- ❑ Costs (BOM)
- ❑ Vendors

- ❑ POC with Vendor/VAR
- ❑ Full scale unit

- ❑ Scale back density?
- ❑ Increase performance?

- ❑ Real customers!

# What do we use it for? Typical use cases.

- ❑ New feature evaluation (e.g., dedup, compression)
- ❑ Protocol evaluation (e.g., NFS v4, SMB3, iSCSI, etc.)
- ❑ Vendor product bakeoffs
- ❑ New architecture / technology evaluation (e.g., SSD caching, tiering)
- ❑ Density testing (how many customers could we deploy on one box?)
- ❑ Routine change management
  - ❑ OS upgrades
  - ❑ Firmware upgrades
- ❑ Etc...

# Storage Engineering team is ...



- ❑ Knowledgeable with storage technologies
- ❑ Eager to investigate and test the leading edge
- ❑ Comfortable with storage protocols and file systems
- ❑ Always challenging hardware and software vendors to stand by their specs
- ❑ Focused on complete picture – cost, performance, capacity, availability



# GoDaddy load generation lab

## Test Application (File, Block, Object storage emulation)



Swift Test Appliance



Used for validating OS/Firmware, configuration updates

Storage Pilot Candidate(s)

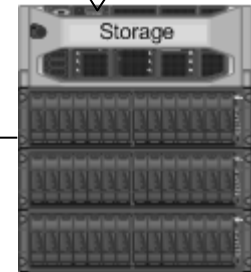


Solution A



Solution B

10Gb switch



Current Production Solution

# How we analyzed our production workload

- ❑ We used vendor provided performance analyzer, netstat, and nfsstat utilities to obtain the workload characteristics.

**For example:**

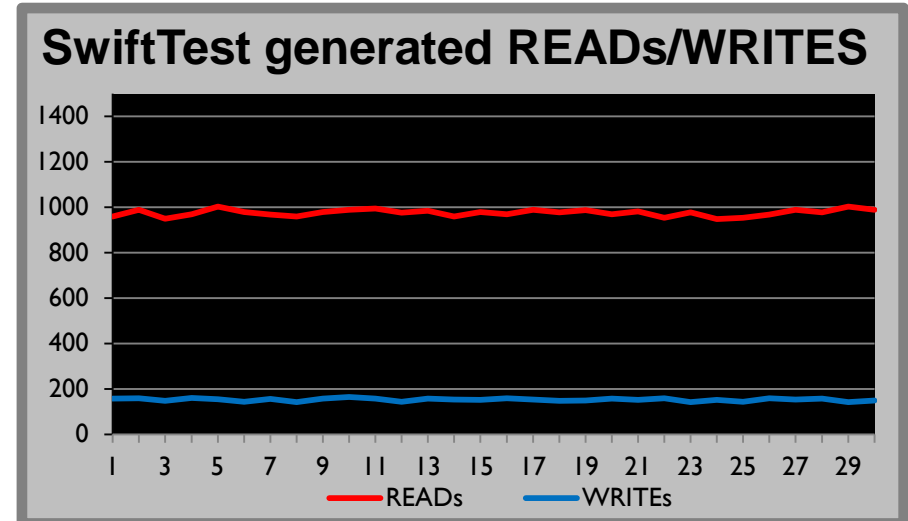
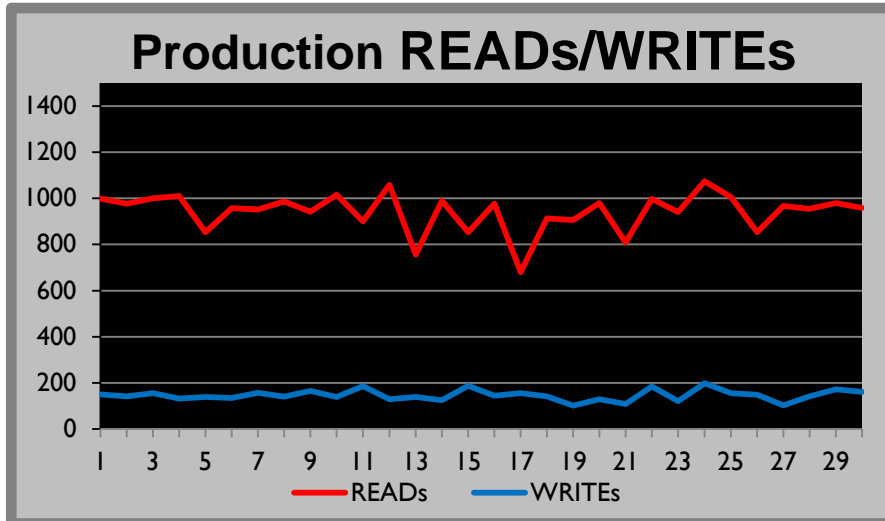
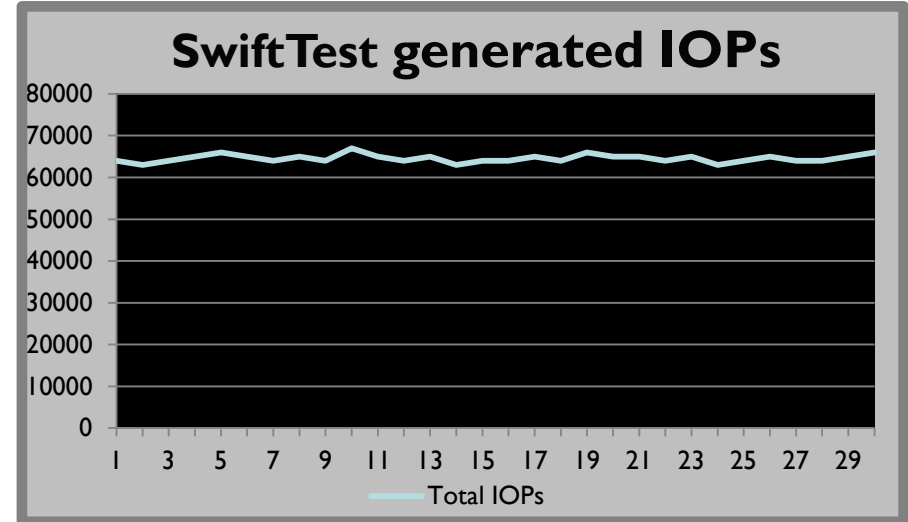
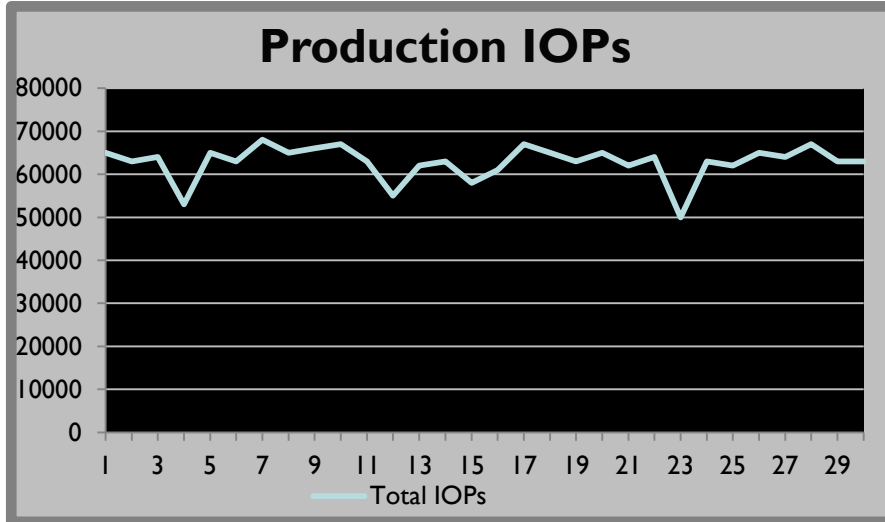
- ❑ Bandwidth utilized
- ❑ NFS Latency
- ❑ READ %
- ❑ WRITE %
- ❑ Metadata %
  - ❑ GETATTR %
  - ❑ LOOKUP %
  - ❑ ACCESS %



# Our production workloads vs. our SwiftTest simulation: virtually identical!

Workload	GoDaddy Production	SwiftTest Simulation
Total NFS ops	~65K	~66K
Avg. Latency	1.5 ms	1.4 ms
• Read	10 ms	11.5 ms
• Write	0.5 ms	0.6 ms
• Other (Metadata Ops)	0.5 ms	0.4 ms
<u>Op-Mix</u>		
• Reads	5%	5%
• Writes	1%	1%
• “Other” Ops	94%	94%
• Getattr	62%	63%
• Lookup	11%	13%
• Access	17%	14%
Avg. CPU Utilization	81%	80%
Max. Disk Utilization	55%	54%

# Actual production workload vs. SwiftTest emulation



# Types of validations conducted

SSDs

Caching /  
Tiering

Compression

New disk /  
OS configs

Commodity  
Hardware

Deduplication



# Validating new hardware configuration

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	<del>30ms</del>	<del>18ms</del>



# Validating caching, tiering and SSDs technologies

	<b>Pass / Fail</b>	<b>Findings</b>
Solution A	PASS	<ul style="list-style-type: none"><li>Cache offloaded ~45% of READS from disks</li><li>Able to test cache “warm-up” time</li></ul>
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 and deal with random reads
Solution D	PASS	Full SSD array was able to push SSDs to required throughput.



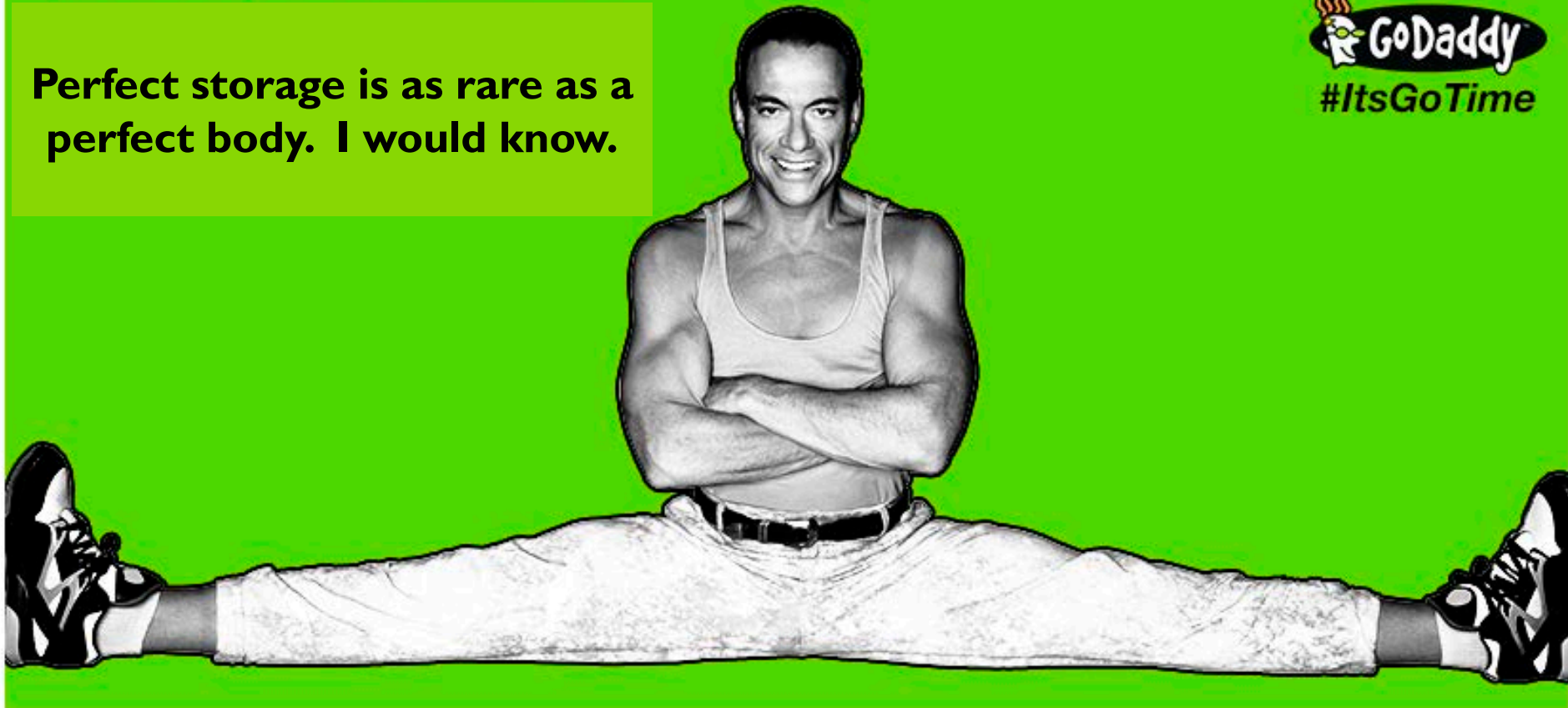
# Validating commodity hardware + open source storage software solutions

	Faster	Cheaper	Better
	Read Latency		
Vendor-A (production reference)	10ms	baseline	+ Features + Support + Mature/stable
Open Compute solution 1	8ms	<b>23% reduction</b>	- Support - Features - Stability unknown
Open Compute solution 2	3ms	<b>56% reduction</b>	- Support - Features - Stability unknown

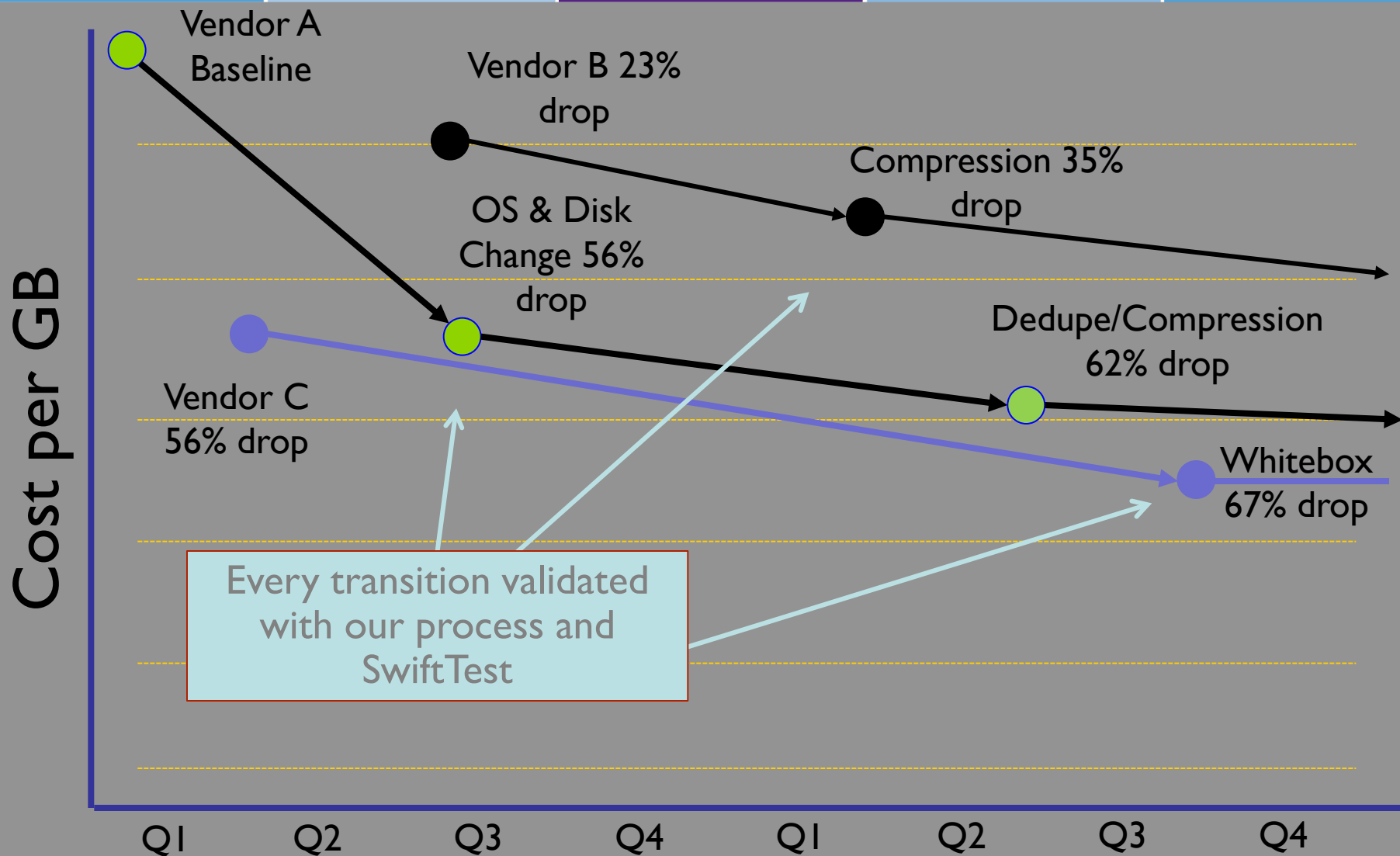




**Perfect storage is as rare as a perfect body. I would know.**



# Benefits: Go Daddy cost reduction - ROI roadmap



# Summary: Benefits of storage validation

**Time to Validate**  
Weeks to test in lab,  
not 12 months in  
production

**Price/performance  
analysis**  
Pick most cost effective  
solution for work load

**Performance  
Testing**  
Confidence in  
user experience

**Freedom to  
Innovate**  
Insight we  
need



**Availability**  
Confidence in 5-9's  
under Load



#ItsGoTime

**Contact us:**

[Julia@godaddy.com](mailto:Julia@godaddy.com)

[jbrichardson@godaddy.com](mailto:jbrichardson@godaddy.com)