Forget IOPS:
A Proper Way to Characterize & Test Storage Performance

Peter Murray
peter@swifttest.com
SwiftTest

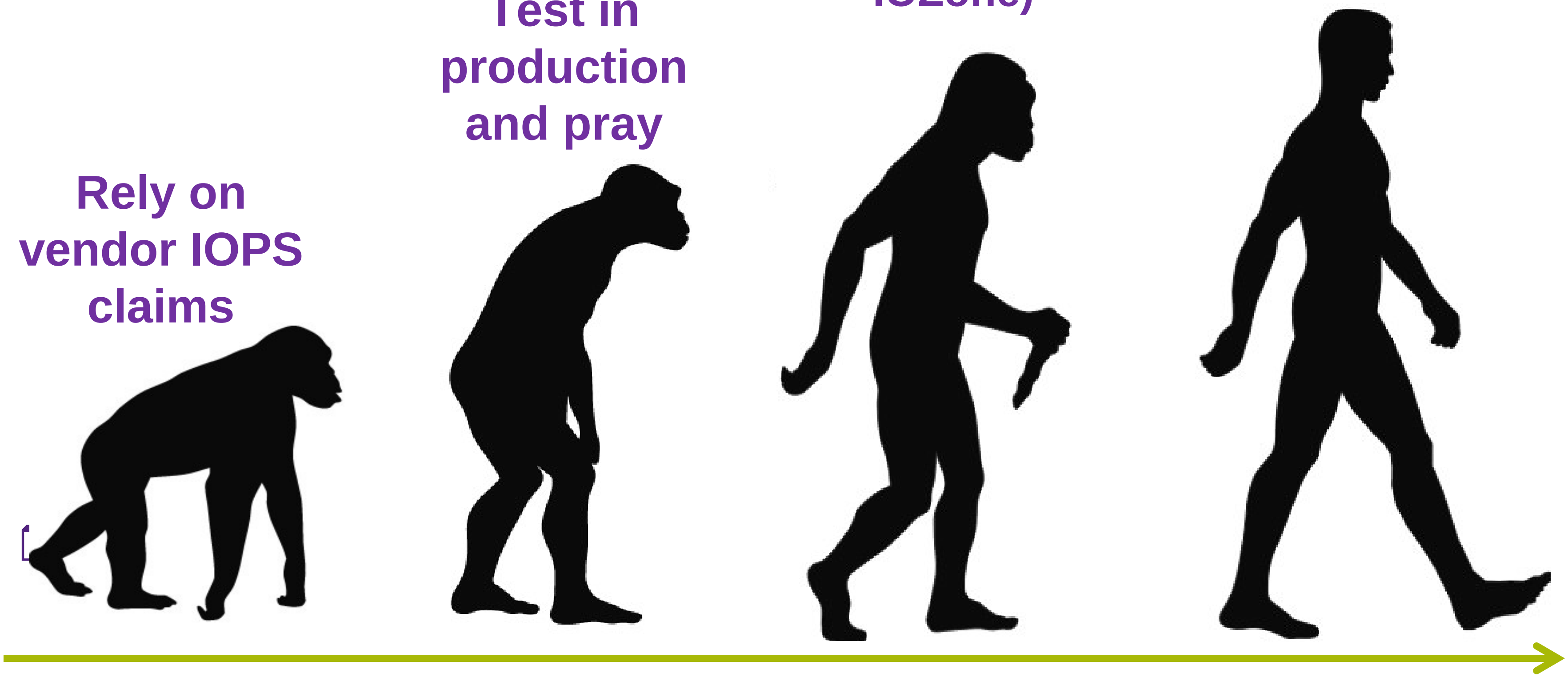
Storage Performance Validation

Rely on vendor IOPS claims

Test in production and pray

Validate with freeware tools (Iometer, IOZone)

Validate with Workload Models



The Evolution of Storage Performance Validation

What is Storage Validation?

FC
SMB
iSCSI
NFS
OBJECT HTTP-S

vmware
Microsoft Exchange Server 2010
ORACLE
Microsoft SQL Server
VDI
SAP

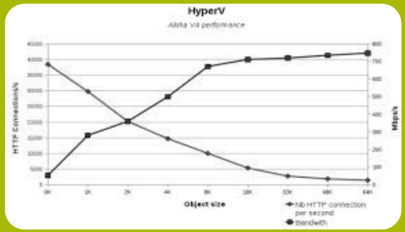
Validation Appliance → **Workload Emulations** → **Storage Under Test**

Storage Infrastructure Performance
Your Customers Care!

Product Evaluations & Vendor Bakeoffs

New Feature & Technology Evaluations

Performance Impact of Virtualization



Optimal Storage Configuration or Protocol

NFS, SMB, FC, iSCSI

Impact of Infrastructure Changes

Troubleshooting

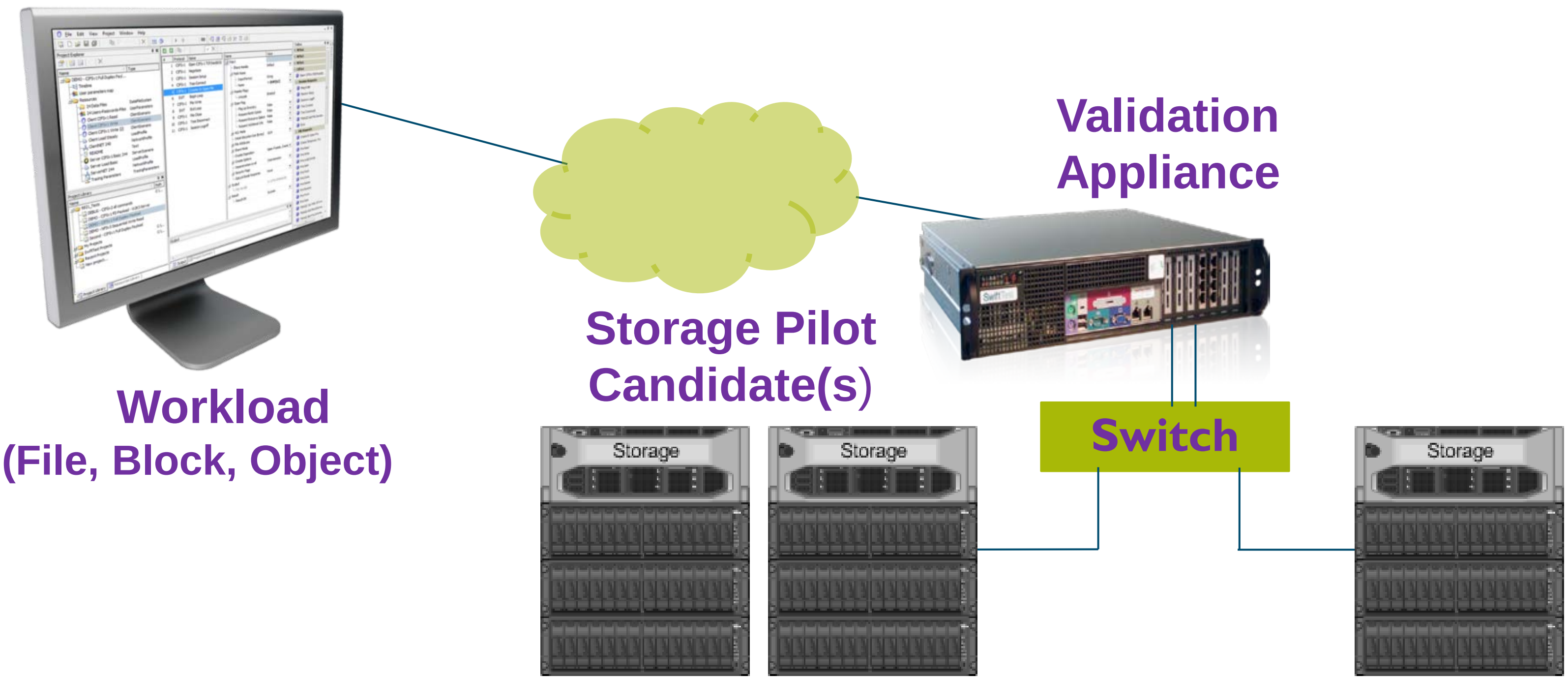
Common Enterprise IT Challenges

- Validate new hybrid SSD architecture & design to reduce storage costs
- Choose most cost effective vendor from short list
- Validate that proposed storage system can meet performance needs
- Configuration planning for private cloud migration
- Troubleshoot poor performance of enterprise BI application

Common Requirements

- Emulate our application workloads with high fidelity
- Test on target system

Validating New Storage Products & Upgrades



Agenda

- What are IOPS?
  - Why IOPS are insufficient for storage testing
- What is metadata?
  - Why metadata is critical for storage testing
- What are workloads?
  - Why workloads are the essential to storage testing

IOPS

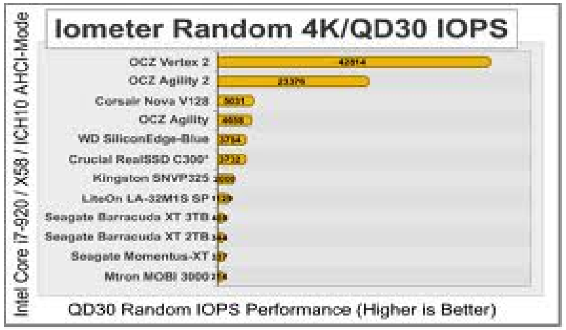
IOPS Definition

- **IOPS** [Storage System]: I/O Operations per Second
  - IOPS/W and MBps/W are the units normally used to report these quantities
  - IOPS and read/write loads are not specified in this definition
  - Care must be taken when comparing different storage systems that the I/O load used is as identical as possible
  - Comparisons of these measurements when I/O sizes and load mix are different may not be very meaningful

From the 2012 SNIA Dictionary. Used with permission. All rights reserved.

Origin of IOPS

- Originally a measurement for local hard disk performance
  - Enabled a common way to test Write/Read disk performance
- Iometer most common IOPS measurement tool
  - Written by Intel, maintained by the Open Source community
  - Designed to test drives, expanded to test NAS performance
  - Widely used to compare overall system performance
  - Commonly cited in vendor literature
  - Uses host’s TCP stack to communicate
  - Designed for Write/Read IO testing
  - Tests include only the metadata required to perform IO



IOPS: One Size Does Not Fit All

- One IOP requires one or more request-response pairs
  - Depends on block/chunk size
    - 1KB – 1 Request, 1 response
    - 4KB – 1 Request, 3 Responses
    - Larger block sizes require more responses
    - Larger block sizes mean a lower IOPS rate
- There is no “standard” size
  - 4KB often cited, but varies – vendor dependent

IOPS Vary in Real World Application Traffic

- Real storage IO does not use a fixed block/chunk size
  - Block/chunk size is protocol version and application dependent
  - Block sizes have increased in newer protocol versions
- Write/Read operations may be a small fraction of storage traffic
  - May comprise less than 10%
  - All remaining traffic is related to metadata
- RDMA Write/Read operations may erode the usefulness of isolated W/R IOPS measurement
  - Less time required for data transfer
  - Metadata still needs to be processed

Metadata

Metadata

- **Metadata operations enable IO**
- **Information about data**
  - Operations needed to access data
    - Location
    - Extent
    - File name
    - Access
    - Attributes, locking, etc.

#	Protocol	Name
1	SMB2	Open SMB2 TCP Connection
2	SMB2	Negotiate
3	SMB2	Session Setup
4	SMB2	Tree Connect
5	SMB2	Create File
6	SMB2	Create File
7	SMB2	Resiliency Request
8	SMB2	Query Info
9	SMB2	File Lock
10	SMB2	File Read
11	SMB2	File Write
12	SMB2	Set Info
13	SMB2	File Close
14	SMB2	Tree Disconnect
15	SMB2	Logoff

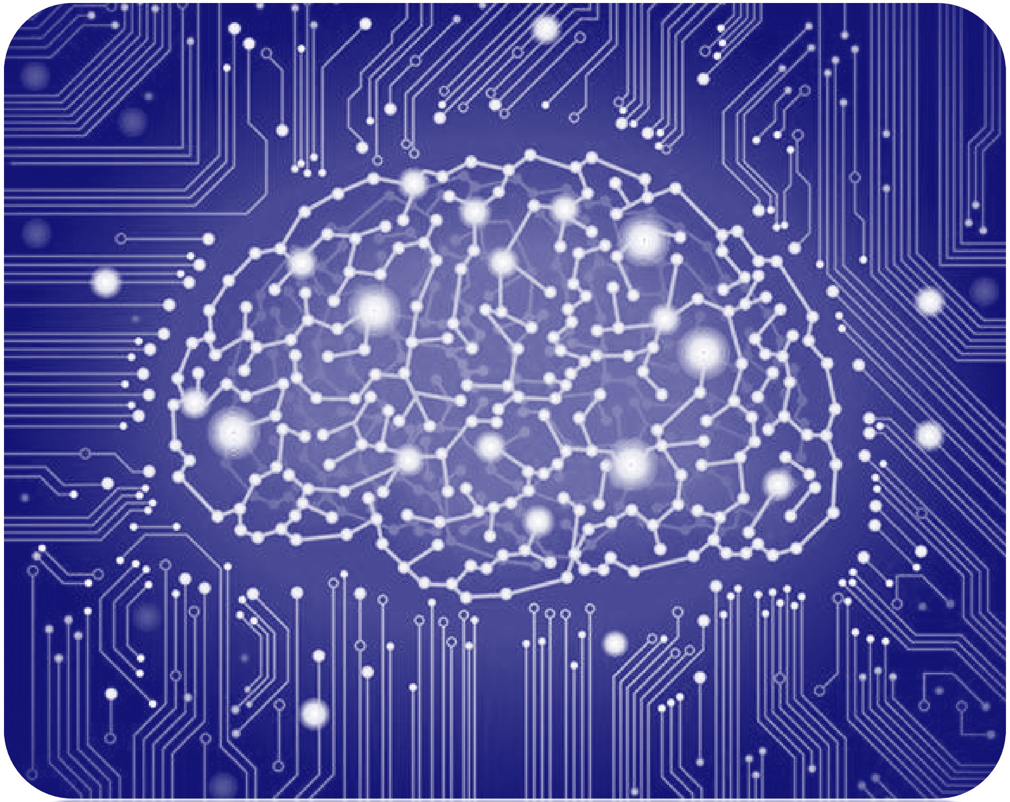
Metadata Usage

- The amount of metadata required to perform IO varies widely
- Determined by access method
  - Often in excess of 50% of storage traffic; sometimes 90%+
  - Block/chunk vs. file
  - Both require metadata
  - File access requires more metadata

Reducing Metadata's Impact

- Metadata impacts performance
  - Based on application type
    - Higher usage: Web apps, Office apps
    - Lower usage: Database, eMail
- Caching with RAM, Flash or SSDs helps, but expensive
  - Metadata performance on HDDs is slow
  - Caching doesn't eliminate Metadata, only speeds processing

Workloads

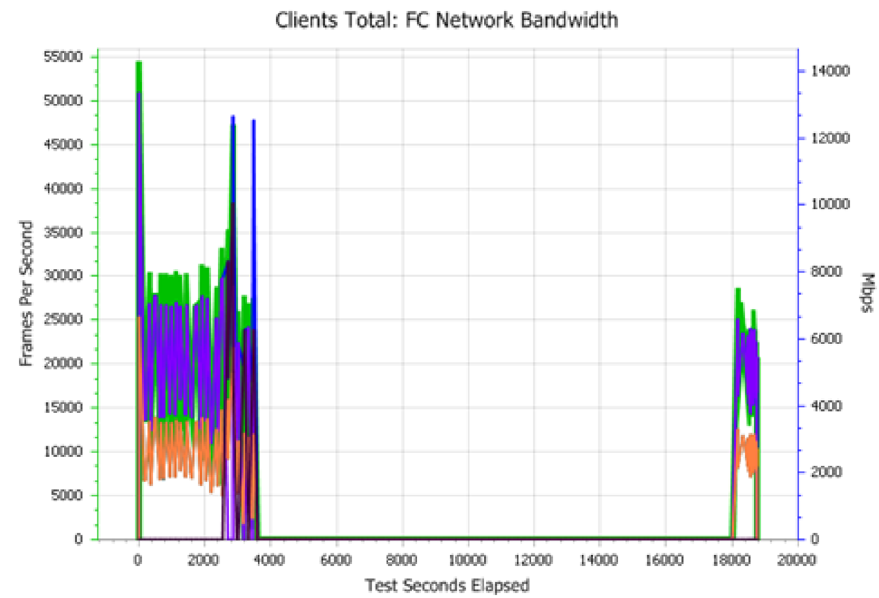


Workloads Defined

- IOPS, metadata and access patterns that reflect an application
- Each application has an unique signature:
  - Write vs. read %
  - Random vs. sequential access
  - IO vs. metadata %
  - Data compressibility
  - Block/chunk size
  - Metadata command frequency
  - Use of asynchronous/compound commands
- Workloads are intrinsic to characterizing storage performance

The Importance of Workloads

- **Workload testing enables engineers to:**
  - Understand how an application operates in a given environment:
    - Server
    - Cluster
    - With Deduplication/Compression
    - With varying network configurations and conditions
- **Understand overload and failure conditions**
  - Degraded application performance vs. application failure
  - Network issues:
    - Active-active network failover
    - Active-passive network failover



Creating & Understanding Workloads

The Workload Model

Access Patterns

File System

I/O Parameters

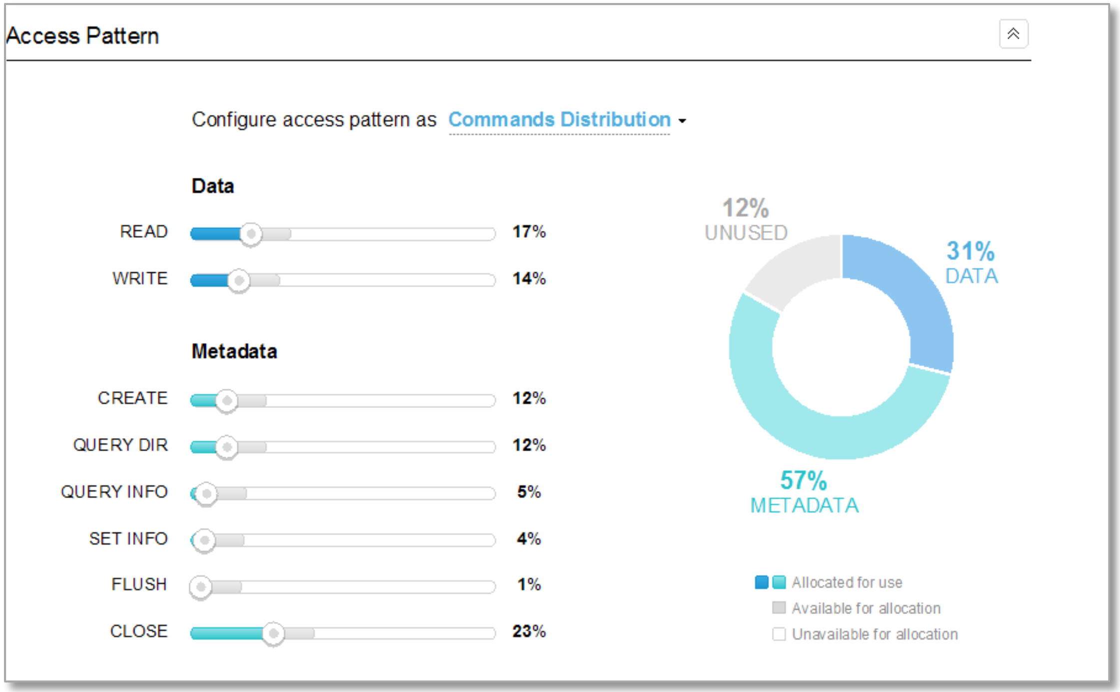
Load Properties

Access Patterns

Access Patterns

I/O Mix

- ❑ Write/Read %
- ❑ Metadata %

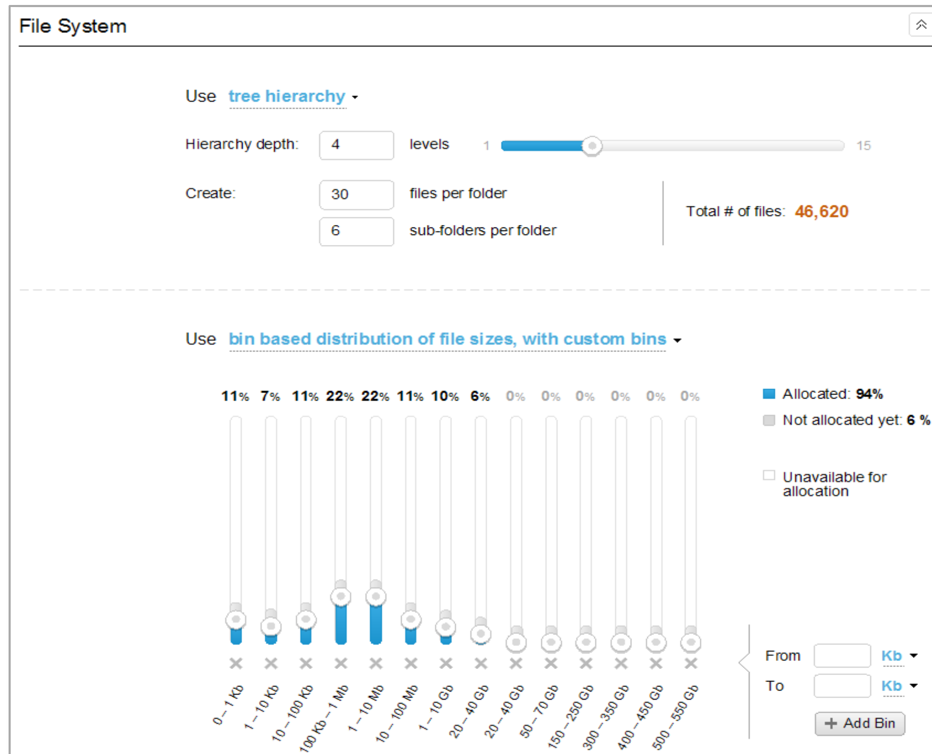


File System

File System

File System Hierarchy

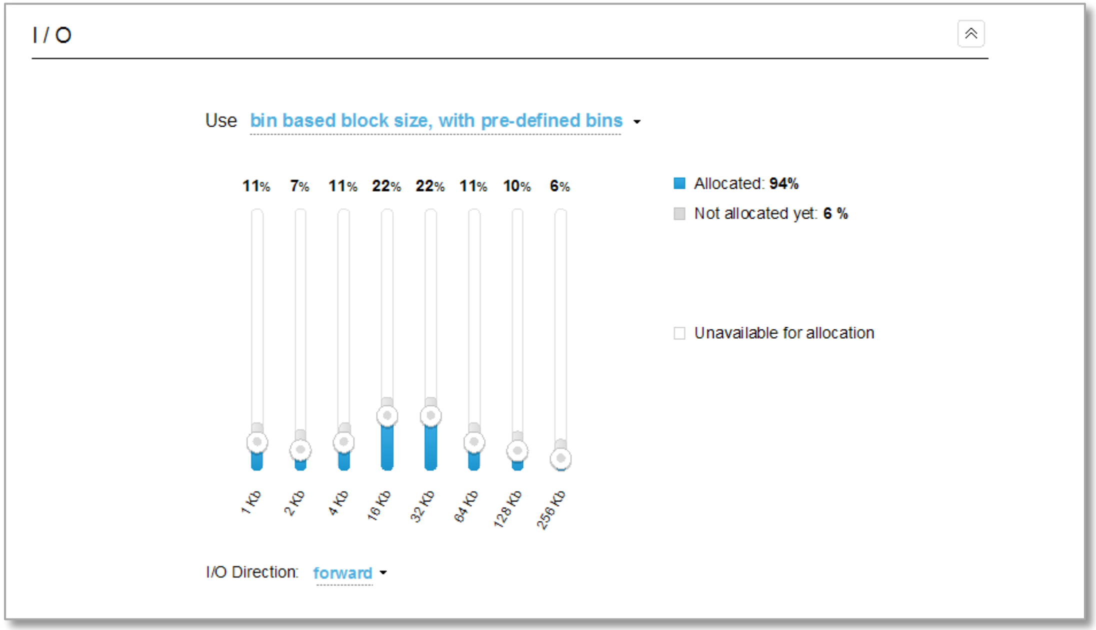
- ❑ File system depth
- ❑ Files per folder
- ❑ Number of subfolders
- ❑ File size distribution



I/O Parameters

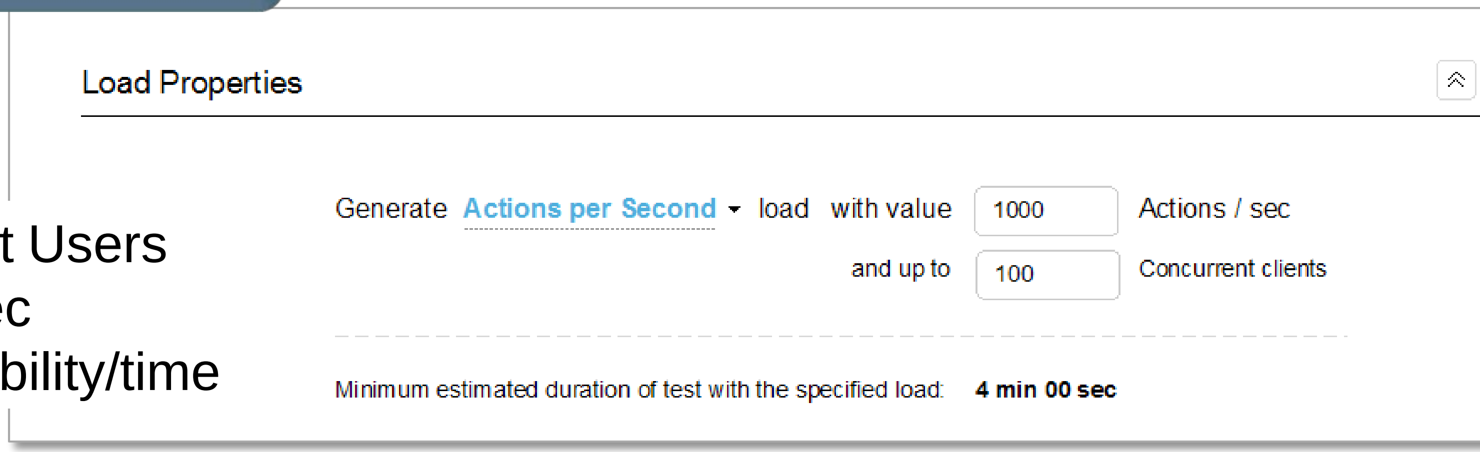
I/O Parameters

- Block/Chunk Size
- Read/Write Direction



Load Properties

Load Properties



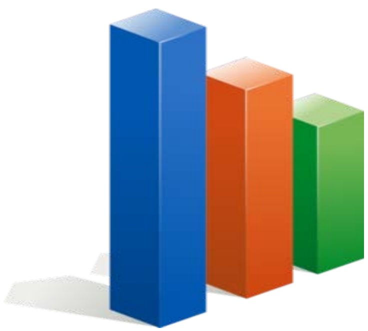
- Concurrent Users
- Actions/Sec
- Load variability/time

Step 1

Creating a Production Workload Model

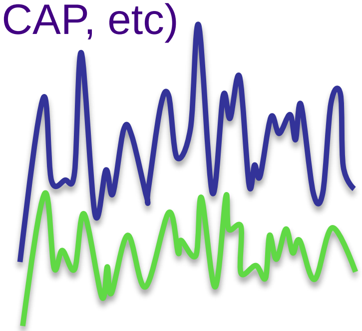
10 YEARS SDC STORAGE DEVELOPER CONFERENCE SNIA SANTA CLARA, 2013

PRODUCTION STATS
(Workload Analyzers, Netstat, NFSstat, etc)



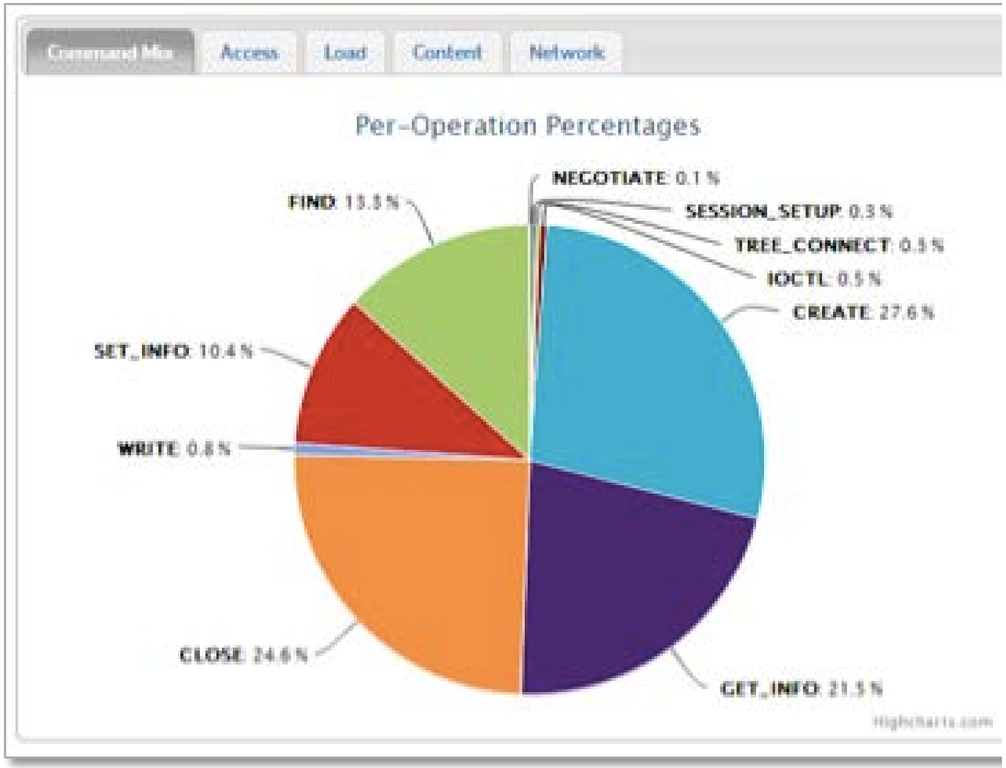
PRE-BUILT TEST SUITES

PACKET CAPTURES
(PCAP, etc)



ACCURATE, REALISTIC WORKLOAD MODEL

Workload Analysis
Command Mix



Representative Command Mix

- **I/O Percentage**
  - Write
  - Read
- **Metadata**
  - Metadata commands
  - Command percentages

Workload Analysis
I/O Mix

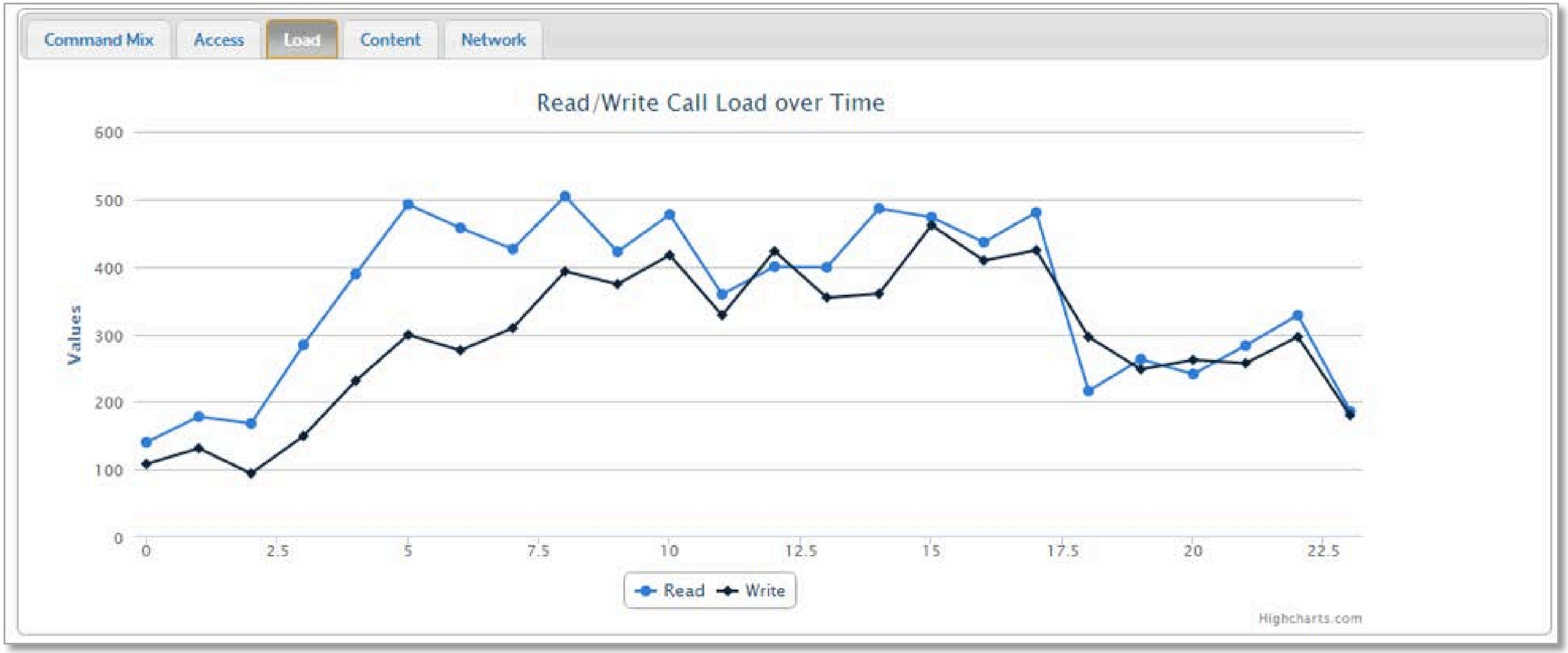
I/O Mix

- Chunk Size
- Read/Write Ratio
- Hot Spots

Load Parameters

Load Parameters

- Temporality
- Burstiness
- Overload conditions



Workload Re-creation Challenges

- Difficult for vendors and operators to vet against real-world service and application conditions at scale:
  - **Big data:** database intensive
  - **Virtualized environments:** VDI, general VM workloads
  - **Infinite permutations**: bare-metal, virtual, Linux, Windows nodes with applications on top in multi-tenant cloud environment
- Server to storage traffic in the data center is a mix of OS, hypervisor and application behaviors
- Complex workload emulation is difficult and time consuming
  - Competency in generating detailed, scalable emulations is highly desirable

Workload Emulation Requirements

- A solution to model different data center conditions –
  - **Flexibility:** allow customers to model a wide spectrum of conditions ranging from exact conversation to blended cloud application mix
  - **Efficiency:** without tedious study, analysis and test configuration synthesis
- A system to analyze/fingerprint traffic in a standardized way
  - Framework that decomposes the characteristics of traffic into quantifiable workload dimensions
- Generate specific or generalized workloads

Workload Simulation Doesn't Require Dozens of Servers / VMs to Drive

	Go Daddy In 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
• 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%

Customer Example: Workloads Modeled from Production

Protocol	File Action Distribution
NFSv3	Getattr 55%
	Lookup 11%
	Access 17%
	Read 5%
	Write 10%

Protocol	File Action Distribution
SMB2	Query path info 27%
	Query file info 25%
	Read 10%
	Write 5%
	Set file info 6%
	NT create 10%
	Close 9%

Protocol	File Size Distribution
NFSv3	8192
	100000
	2000000

Protocol	File Size Distribution
SMB2	4096
	100000
	1000000
	10000000

Step 2

Run Workload Model Against Target

FC

SMB

iSCSI

NFS

OBJECT
HTTP-S

vmware

Microsoft Exchange Server 2010

ORACLE

Microsoft SQL Server

VDI

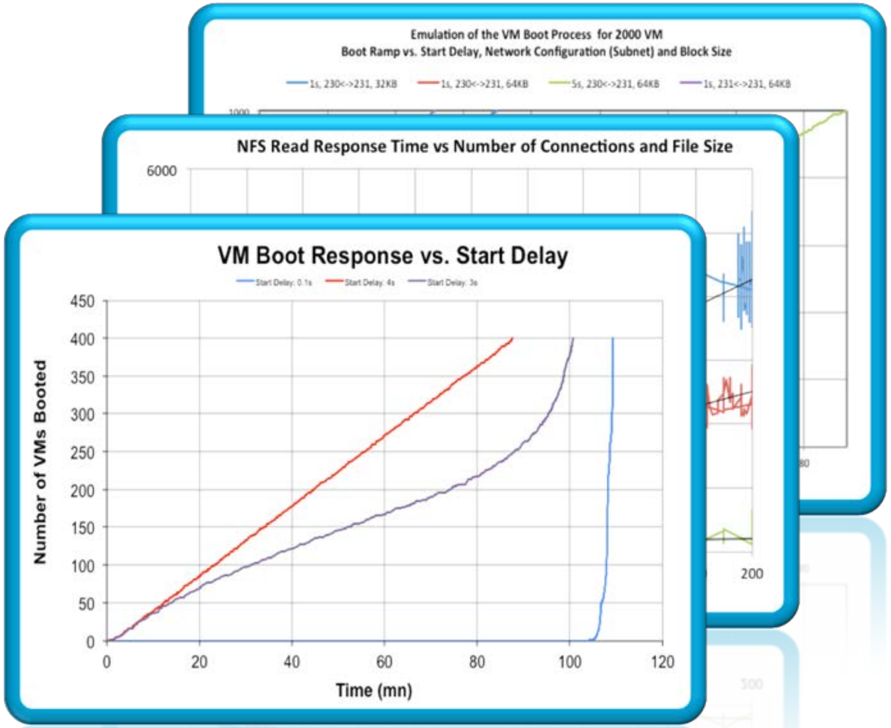
SAP

Validation Appliance   Workload Emulations   Storage Under Test

Step 3

Analyze Results for Better Decisions

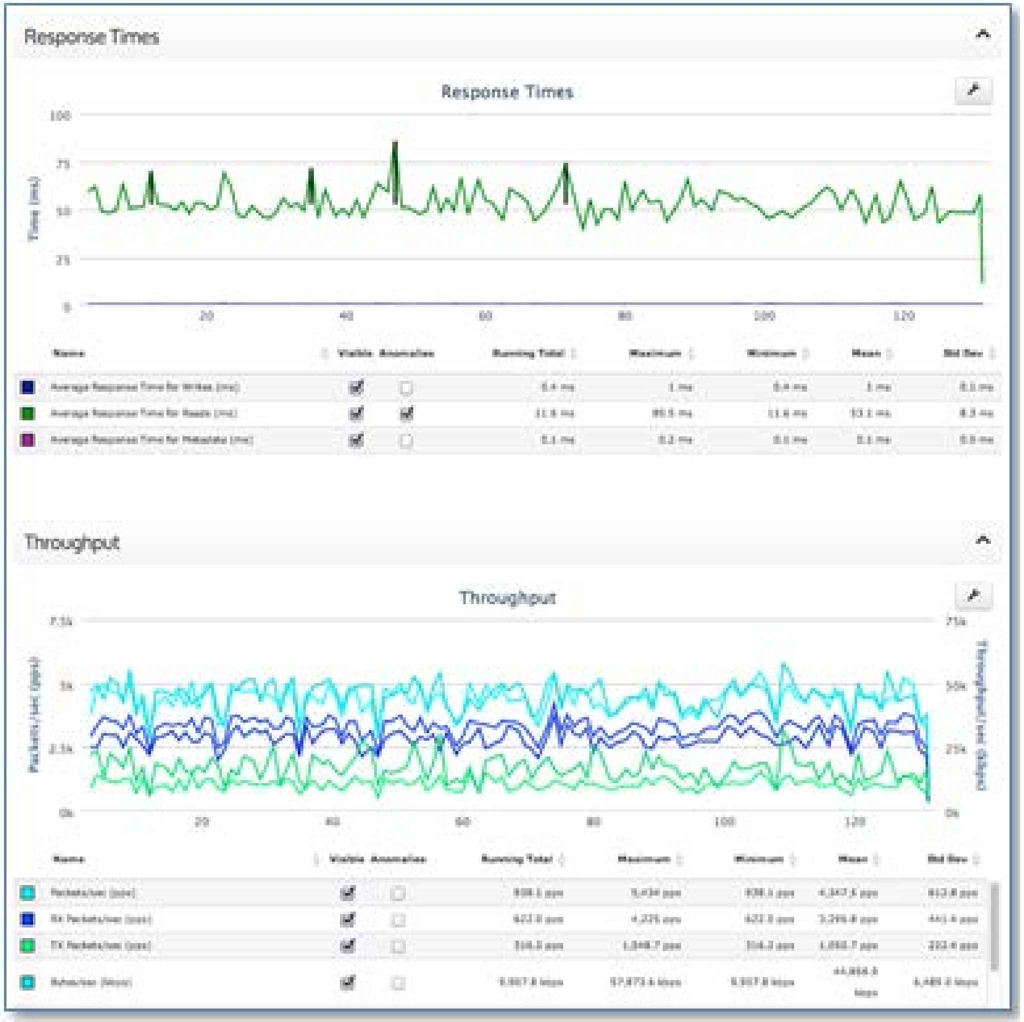
Analytics



Insight

- Block vs File
- Boot storm handling
- Limits testing
- Failure modes
- Effects of flash, dedupe, tiering, scale-out, etc.

Results Analysis
Performance



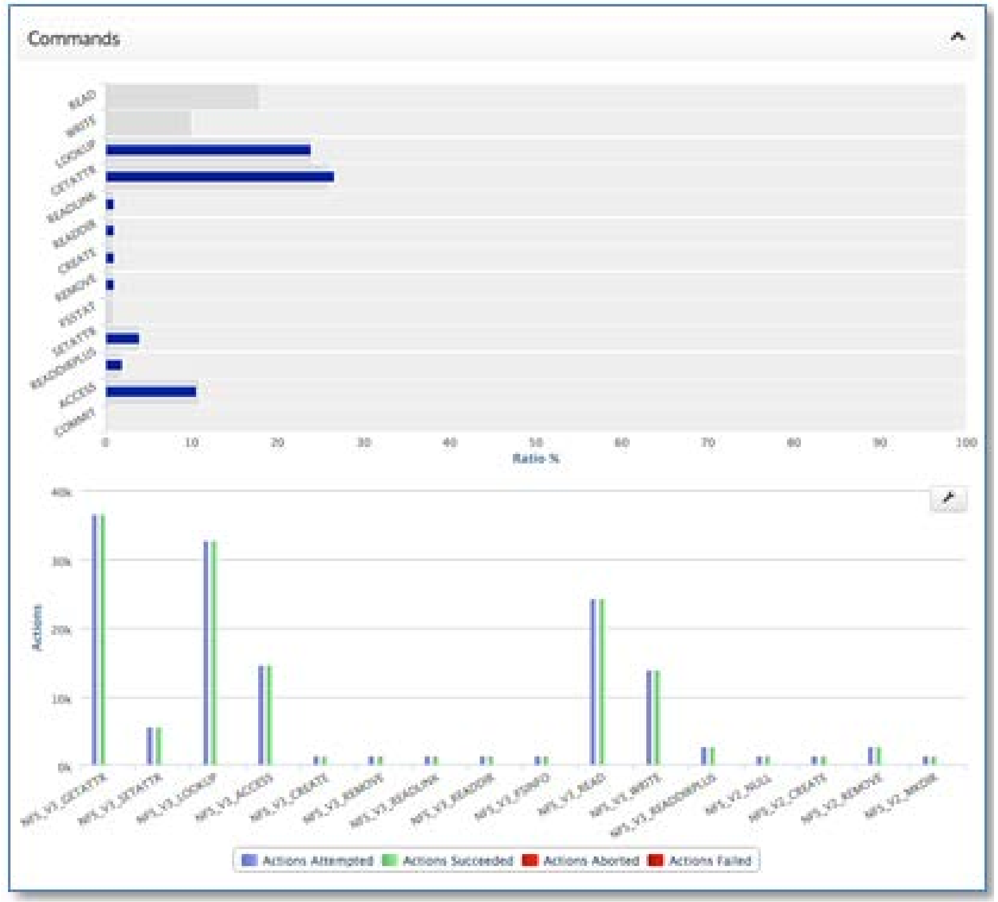
Performance

- Response times
- Throughput

Results Analysis
Command Mix

Expected Command Mix

- Execution status
- Attempts
- Successes
- Errors
- Aborts



Summary

- IOPS measurement alone cannot characterize real application storage performance
- Inclusion of metadata is essential
- Workload modeling and purpose-build load generation appliances are the way to emulate applications
- The more complete the emulation, the deeper the understanding