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

Peter Murray
peter@swifttest.com
SwiftTest
Storage Performance Validation

The Evolution of Storage Performance Validation

1. Rely on vendor IOPS claims
2. Test in production and pray
3. Validate with freeware tools (Iometer, IOZone)
4. Validate with Workload Models

- Validate with Workload Models
- Validate with freeware tools (Iometer, IOZone)
- Test in production and pray
- Rely on vendor IOPS claims
What is Storage Validation?
Storage Infrastructure Performance
Your Customers Care!

- Product Evaluations & Vendor Bakeoffs
- New Feature & Technology Evaluations
- Performance Impact of Virtualization

- Optimal Storage Configuration or Protocol
- Impact of Infrastructure Changes
- Troubleshooting

NFS, SMB, FC, iSCSI
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

Workload (File, Block, Object)

Storage Pilot Candidate(s)

Validation Appliance

Switch

Solution A

Solution B

Current Production Solution
(to validate OS/Firmware upgrades, configuration changes)
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.

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

I/O Mix

- Write/Read %
- Metadata %
File System

File System Hierarchy

- File system depth
- Files per folder
- Number of subfolders
- File size distribution
I/O Parameters

- Block/Chunk Size
- Read/Write Direction
Load Properties

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

Generate Actions per Second - load with value 1000 Actions / sec and up to 100 Concurrent clients

Minimum estimated duration of test with the specified load: 4 min 00 sec
Step 1

Creating a Production Workload Model

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

PRE-BUILT TEST SUITES

PACKET CAPTURES
(PCAP, etc)

ACCURATE, REALISTIC WORKLOAD MODEL

2013 Storage Developer Conference. © Insert Your Company Name. All Rights Reserved.
Workload Analysis
Command Mix

Representative Command Mix

- I/O Percentage
  - Write
  - Read
- Metadata
  - Metadata commands
  - Command percentages
Workload Analysis

I/O Mix

- Chunk Size
- Read/Write Ratio
- Hot Spots
Workload Analysis

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

<table>
<thead>
<tr>
<th></th>
<th>Go Daddy In Production</th>
<th>SwiftTest Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total NFS ops</td>
<td>~65K</td>
<td>~66K</td>
</tr>
<tr>
<td>Avg. Latency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Read</td>
<td>1.5 ms</td>
<td>1.4 ms</td>
</tr>
<tr>
<td>- Write</td>
<td>10 ms</td>
<td>11.5 ms</td>
</tr>
<tr>
<td>- Metadata Ops</td>
<td>0.5 ms</td>
<td>0.6 ms</td>
</tr>
<tr>
<td>Avg. CPU Utilization</td>
<td>81%</td>
<td>80%</td>
</tr>
<tr>
<td>Max. Disk Utilization</td>
<td>55%</td>
<td>54%</td>
</tr>
<tr>
<td>Avg. Latency</td>
<td>1.5 ms</td>
<td>1.4 ms</td>
</tr>
<tr>
<td>- Read</td>
<td>10 ms</td>
<td>11.5 ms</td>
</tr>
<tr>
<td>- Write</td>
<td>0.5 ms</td>
<td>0.6 ms</td>
</tr>
<tr>
<td>- Metadata Ops</td>
<td>94%</td>
<td>94%</td>
</tr>
<tr>
<td>Op-Mix Metdata Ops</td>
<td>62%</td>
<td>63%</td>
</tr>
<tr>
<td></td>
<td>11%</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>17%</td>
<td>14%</td>
</tr>
</tbody>
</table>

Go Daddy.com
## Customer Example: Workloads Modeled from Production

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NFSv3</td>
<td>Getattr 55%</td>
<td>SMB2</td>
<td>Query path info 27%</td>
</tr>
<tr>
<td></td>
<td>Lookup 11%</td>
<td></td>
<td>Query file info 25%</td>
</tr>
<tr>
<td></td>
<td>Access 17%</td>
<td></td>
<td>Read 10%</td>
</tr>
<tr>
<td></td>
<td>Read 5%</td>
<td></td>
<td>Write 5%</td>
</tr>
<tr>
<td></td>
<td>Write 10%</td>
<td></td>
<td>Set file info 6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NT create 10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Close 9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protocol</th>
<th>File Size Distribution</th>
<th>Protocol</th>
<th>File Size Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFSv3</td>
<td>8192</td>
<td>SMB2</td>
<td>4096</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td></td>
<td>100000</td>
</tr>
<tr>
<td></td>
<td>2000000</td>
<td></td>
<td>10000000</td>
</tr>
</tbody>
</table>
Run Workload Model Against Target

Step 2

Validation Appliance  Workload Emulations  Storage Under Test

FC  SMB  iSCSI  NFS  OBJECT  HTTP-S

vmware®

Microsoft® Exchange Server 2010

ORACLE®

Microsoft® SQL Server

VDI  SAP®

2013 Storage Developer Conference. © Insert Your Company Name. All Rights Reserved.
Step 3

Analyze Results for Better Decisions

Analytics

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

Insight
Results Analysis

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