Storage Performance and IO Load Basics

Leah Schoeb, Vice Chair, SNIA Technical Council

SNIA Emerald™ Training
SNIA Emerald Power Efficiency Measurement Specification, for use in EPA ENERGY STAR®

June 24-27, 2013
Topics

- Today’s Impact on Storage Performance
- Storage Performance Planning
- Troubleshooting Methodology and basic metrics
IO Performance Needs Monitoring at Every Level

**Application Level**
App Specific Perf tools/stats

**Guest OS**
CPU Utilization, Memory Utilization, I/O Latency

**Virtualization Level**
Performance Metrics /Charts Limits, Shares, Virtualization Contention

**Physical Server Level**
CPU and Memory Saturation, Power Saving

**Connectivity Level**
Network/FC Switches and data paths Packet loss, Bandwidth Utilization

**Storage Level**
SAN or NAS Devices Utilization, Latency, Throughput
Storage Performance Planning
Planning for Performance

Storage Planning → Workload Behavior → Storage Optimization
Storage Planning

- Understand the workload
- Sharing or Consolidation
- Storage Protocol Options
  - File, block, or object
- Data Reduction Options
  - Thin provisioning
- Data Protection
- Other Storage Technology trade-offs
## Rotating Media Selection

### RAID Levels
- **RAID 0**: $\quad$
- **RAID 5**: $$
- **RAID 6**: $$$$
- **RAID 10**: $$$$$$

### Drive Types and Specifications

<table>
<thead>
<tr>
<th>Drive Type</th>
<th>Speed</th>
<th>MB/sec</th>
<th>IOPS</th>
<th>Latency</th>
<th>LC Manage</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC 4Gb</td>
<td>15k</td>
<td>150</td>
<td>200</td>
<td>5.5ms</td>
<td>High Perf. Trans</td>
</tr>
<tr>
<td>FC 4Gb</td>
<td>10k</td>
<td>75</td>
<td>165</td>
<td>6.8ms</td>
<td>High Perf. Trans</td>
</tr>
<tr>
<td>SAS (6Gb, 12Gb)</td>
<td>10k</td>
<td>150</td>
<td>185</td>
<td>12.7ms</td>
<td>Streaming</td>
</tr>
<tr>
<td>SATA (6Gb, 12Gb)</td>
<td>7200</td>
<td>140</td>
<td>38</td>
<td>12.7ms</td>
<td>Streaming/Nearline</td>
</tr>
<tr>
<td>SATA</td>
<td>7200</td>
<td>68</td>
<td>38</td>
<td>12.7ms</td>
<td>Nearline</td>
</tr>
</tbody>
</table>
Solid State Storage

- No all SSDs designed the same
  - NAND-based flash memory
  - DRAM-based (Random Access Memory)
  - Enterprise flash drives (EFDs)
  - Hybrid Drives

- Performance varies widely
  - Capacity
  - Compression
  - Wear leveling
  - Error Correction and bad block mapping
  - Metadata management
  - Garbage collection
  - Encryption
# Solid State Storage

<table>
<thead>
<tr>
<th>Metric</th>
<th>SLC</th>
<th>MLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency (microseconds)</td>
<td>100</td>
<td>200-300</td>
</tr>
<tr>
<td>Persistence</td>
<td>10x more persistent</td>
<td>Less reliable*</td>
</tr>
<tr>
<td>Cost</td>
<td>30% more expensive</td>
<td>More cost effective</td>
</tr>
<tr>
<td>Sequential read/writes</td>
<td>3x faster</td>
<td>Slower</td>
</tr>
</tbody>
</table>

*This can be overcome, even reversed by the internal design using higher over provisioning, interleaving, and changes to writing algorithms.
Virtualize to consolidate

Physical

Virtual

SQL
Win2k3
5 Disks

SQL
Win2k3
5 Disks

SQL
Win2k3
5 Disks

APP
OS

vSphere

ESXi Server

ESXi Server

VMDK

VMDK

VMDK
Over Provisioning

- Using Thick provisioning it is easy to over provision.
- You may want to consider Thin Provisioning.
- Most vendors offer Thin Provisioning.
Planning for Performance

- Storage Planning
- Workload Behavior
- Storage Optimization
As I/O workloads increase so does the Response Time

Acceptable Response Time Threshold

Meeting SLA

Not Meeting SLA
Seasonal/Periodic Performance Surges

Seasonal Workload Surges

Normal Workload Activity

Meeting SLA

Not Meeting SLA
Single vs. Multi-threaded Applications

Single Threaded

Multi Threaded

= 8 ms

= 1 ms
I/O Queue Depth

- The number of I/O request waiting to be completed
  - Also known as outstanding I/Os
- Limiting host I/O demands
- Certain applications, under extreme load, can gain performance by increasing the I/O Queue Depth
- Accepting requests from the Application
### Skew

- Asymmetry of a distribution about its mean or the non-uniform distribution of data or I/O activity across storage devices.
- New storage technologies are handling this automatically
- Disk skew
  - An area of the disk has higher amounts of activity
  - Referred to as a ‘hot spot’
  - Data is accessed more frequently
- Controller skew
  - A controller has a higher amount of activity compared to rest of the controllers in a storage system.
Misalignment

Before Partition Alignment

Cluster VMDK (NTFS)  Cluster VMDK (NTFS)  Cluster VMDK (NTFS)  Cluster VMDK (NTFS)

Block (VMFS)  Block (VMFS)  Block (VMFS)

Chunk (SAN)  Chunk (SAN)  Chunk (SAN)

After Partition Alignment

Cluster VMDK (NTFS)  Cluster VMDK (NTFS)  Cluster VMDK (NTFS)  Cluster VMDK (NTFS)

Block (VMFS)  Block (VMFS)  Block (VMFS)

Chunk (SAN)  Chunk (SAN)  Chunk (SAN)
Workload Consolidation

Group similar workloads together (Random w/ Random and Sequential /w Sequential)

Too many sequential threads on a lun will appear as a random workload to the storage
Negative Impact on Sequential Perf.

Mixing Sequential with Random can hurt Sequential workload throughput.
Negative Impact on Sequential Perf.
Mixed Workloads

- OLTP
- Video Streaming
- Email
- Satellite Streaming
- Log Streaming
Planning and Best Practices

Storage Planning → Workload Behavior → Storage Optimization
Optimizing Storage

Over 80% of storage related performance problems stem from misconfigured storage hardware

- Consult SAN Configuration Best Practice Guides
- Ensure disks are correctly distributed
- Ensure the appropriate controller cache is enabled
- Count the cost in choosing a level of protection
Optimizing Storage

- Avoid negatively impacting high volume sequential performance
- Choose a storage protocol best fitting requirements and needs
- Use the Hypervisor filesystem (VMFS, ZFS, SMB3, etc…)
  - No overhead compared to RDM (physical or virtual)
- Thick provisioning
  - Use when possible to help prevent over provisioning
  - No performance impact compared to Thick
- Are other departments sharing a RAID set
Troubleshooting Methodology
Storage Performance

101 BASICS
Performance Methodology

Modeling
Get to know your workload

Measure & Monitor
Use Tools

Performance Analysis
Performance Improve

Validate Success

SNIA Emerald™ Training ~ June 24-27, 2013
www.sniaemerald.com
Understanding Your Workload

- **Workload Indicators**
  - Demand for resources vs. Resources currently used
  - Result is a percentage of Workload
    - Low latency number is Good – Object has the resources it needs
    - Can go above 100% - Object is “Starving”

- **Workload summarized across critical resources**

- **Workload Details View**
  - Detailed understanding of the lacking resource and associated metrics
  - View the state of the Peer and Parent Objects and troubleshoot
    - Am I a victim or a villain?
    - Is this a population problem?
    - Should we move the VM?
    - A Configuration issue?
    - Lack of resources?
Understanding Your Workload

Server Hourly Utilization

- CPU
- Disk I/O
- Network I/O
- Memory

Analyze all resource dimensions
Performance Methodology

Modeling
Get to know your workload

Measure & Monitor
Use Tools

Performance Analysis
Performance Improve

Validate Success

SNIA Emerald™ Training ~ June 24-27, 2013
www.sniaemerald.com
Approach to Real-Time Performance Management

3rd Generation – Holistic, Real Time Analytics

- Flexible INTEGRATION to many data sources
- Enterprise SCALABILITY
- Patented performance ANALYTICS
- Powerful information DASHBOARDS

I can put all my monitoring tools to good use and get better performance analytics.

Patented performance analysis formula:

$$\sigma_{k,k-1}^2 = \frac{1}{k-2} \sum_{i=1}^{k-1} w_i^2 - \frac{k-1}{k-2} \overline{w}_{k-1}^2$$

SNIA Emerald™ Training ~ June 24-27, 2013

www.sniaemerald.com
Infrastructure vs. Operations Impacts on the storage performance & efficiency

**Performance**
- App = host
- Limited movement

**Capacity**
- Dedicated resources
- Pre-committed

**Configuration**
- Static, pre-configured
- 1-1 mapping

---

**Operational/Cloud OS**
- Setup
- Update
- Inventory
- HA
- vMotion
- DRS
- Distr. S/W
- I/O Control

---

**INFRASTRUCTURE**
- Physical Datacenters

**OPERATIONS**
- Virtual Datacenters
Performance Methodology

Modeling
Get to know your workload

Measure & Monitor
Use Tools

Performance Analysis
Performance Improve

Validate Success
Basic Metrics

- **Performance (Data at work)** – I/O per second (IOPS)
- **Throughput (Data on the move)** - Mega- or Giga- bytes per second (MB/sec, GB/sec)
  - Network throughput Mega- or Giga- bits per second (Mbps, Gbps)
- **Idle (Data at rest)**
- **Response time**
  - HHDs – milliseconds (ms)
  - SSS – microseconds
  - Overall response times – milliseconds (ms)
- **Retries**
- **Queue Depth**
Basic Metrics

➢ Power performance - I/Os per watt
➢ Write coalescing
  ✓ Combining several or many small blocks into one large block then writing that single large block to disk
➢ Hard Disk Drive Service Time:
  ✓ Seek - The initial operation a disk performs to place the read/write head on the right track of a disk drive.
  ✓ Latency (Rotational Latency) - The secondary operation that occurs after the “seek”, which is the time it takes for the data to reach the read/write head of a disk drive.
  ✓ Transfer Time – The time it takes for data to be read from or written to the host after seek and latency.
  ✓ Service Time = seek + latency + transfer Time
### Identifying Unhealthy Storage

<table>
<thead>
<tr>
<th>Metric</th>
<th>Described</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Device latency</td>
<td>Latencies from the storage system</td>
<td>10-15 ms</td>
</tr>
<tr>
<td>Average Kernel latency</td>
<td>Latencies from the kernel’s I/O subsystem</td>
<td>1-2 ms</td>
</tr>
<tr>
<td>Aborts and retries</td>
<td>Can’t keep up with demand and times out or something broke</td>
<td>1</td>
</tr>
<tr>
<td>Response Time</td>
<td>Overall application or OS response time</td>
<td>Many IOs above 10 ms</td>
</tr>
</tbody>
</table>
Performance Methodology

Modeling
Get to know your workload

Measure & Monitor
Use Tools

Performance Analysis
Performance Improve

Validate Success

Improve & Validate

SNIA Emerald™ Training ~ June 24-27, 2013
www.sniaemerald.com
Monitor and Validate Success

- Does your application continue meet its SLA?
- Do known activities perform the same or better?
- Check and monitor key performance counters
- Are business and application owners satisfied?
I/O Generator Tools

101 BASICS
I/O Generators - IOmeter

- **I/O disk testing tool**
  - Uniform distributions (speeds and feeds) ONLY
  - Built originally to measure server side disk storage

- **IOmeter was formerly known as “Intel's Galileo”**.

- **IOmeter does for a computer’s I/O subsystem what a dynamometer does for an engine (Block only)**
  - It measures performance under a controlled load.

- **Measures**
  - Performance and throughput of disk and network controllers.
  - Bandwidth and latency capabilities of buses.
  - Shared bus performance.
  - System-level hard drive and network performance.
I/O Generators - IOMeter

- An access pattern contains mainly the following parameters:
  - **Transfer Request Size** - a minimal data unit to which the test can apply.
  - **Percent Random/Sequential Distribution** - percentage of random requests (read/write ratio)
  - **Percent Read/Write Distribution** - percentage of requests for reading.
  - **# of Outstanding I/Os** - defines a number of simultaneous I/O requests for the given worker and, correspondingly, disc load.
I/O Generators - Vdbench

- I/O workload generator
  - Both uniform and non-uniform distributions
  - Built to measure storage systems
- Generates and measure storage performance (block or file)
- Collect and replay real world enterprise application workloads with the addition of SWAT
- Swiss army knife of I/O generators
- Java based is ported to most major operating systems
  - Unix, Linux, windows, etc…
I/O Generators - Summary

- Many IO Generators
- Uniform vs. non-uniform distributions
- Skew
- Replay real world workloads
- Measuring a disk vs. a storage system
- Measuring block vs file
Thank You

Leah Schoeb
leah@evaluatorexportgroup.com
Twitter: @vLeahSchoeb