SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
Storage Performance Management (SPM), Storage Architecture and SMI-S

This session will appeal to Storage Managers, Performance and Capacity Managers, and those that are seeking a fundamental understanding of storage performance management. This session includes an overview of the processes, technology and skills required to implement SPM, as well as an overview of disk storage system architecture, and the SMI-S specification as it relates to block level performance. The focus is on block level storage systems.
Storage Performance Management (SPM)
How Large is Storage Expenditure?

Where will you be in 2012?

Reduce storage costs by implementing SPM

IT Storage Spend Forecast (Billion)

Source: [1]
What is Storage Performance Management (SPM)?

Processes to ensure that applications receive the required service levels from storage systems, while storage assets are efficiently used.

This means ensuring that the storage hardware resources are used **efficiently** with neither unnecessary hardware components nor components reaching critical utilization levels! It is not the same as Storage Resource Management (SRM) as SRM is about space management.
Benefits of SPM

Three Cases for Reducing Storage costs

- **Risk reduction**: Loss of business by performance outage that could have been avoided
- **Right Sizing**: Picking the exact right storage options results in more value for less money
- **Storage optimization**: hardware savings

How about being able to delay a 40 million USD investment for 9 months?
Risks

The cost difference between these two is often **30% or more** of all storage.

- **Under configured:** Performance issues
- **Sweet spot:** Right performance, Right Cost
- **Over configured:** Efficiency issues

Risk is **not** removed: Hot spots will still occur from time due to imbalance.

- Higher
  - Performance
  - Cost
  - Utilization levels

→ size (and cost) of storage configuration
# SPM Maturity Stages

<table>
<thead>
<tr>
<th>SPM Maturity Stages</th>
<th>Reactive</th>
<th>Proactive</th>
<th>Predictive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem Identification</strong></td>
<td>End-user identified</td>
<td>Automatic with early warning</td>
<td>Growth modeling</td>
</tr>
<tr>
<td><strong>Problem Resolution</strong></td>
<td>Lengthy, may require vendor</td>
<td>Quick, vendor independent</td>
<td>Future focused</td>
</tr>
<tr>
<td><strong>Storage Sizing</strong></td>
<td>Vendor configured</td>
<td>Estimate from historical data</td>
<td>Model exact configurations</td>
</tr>
<tr>
<td><strong>Volume Placement</strong></td>
<td>Randomly placed</td>
<td>Manual optimization</td>
<td>Intelligent balancing</td>
</tr>
<tr>
<td><strong>Service Level Attainment</strong></td>
<td>Lowest</td>
<td>High</td>
<td>Highest</td>
</tr>
<tr>
<td><strong>Storage Hardware Costs</strong></td>
<td>Highest</td>
<td>Low</td>
<td>Lowest</td>
</tr>
</tbody>
</table>

---

Storage Performance Management
© 2012 Storage Networking Industry Association. All Rights Reserved.
Stakeholders

Better Quality / Lower Cost

Trending/Space

IT

SMF Focal

Evaluate Technology

Data Placement, Data Collection

Resolve Problems

Architecture

Administrator

Unix

Windows

MF

Storage

Server

Perf/Cap
Trends Affecting Storage Performance
HDD Capacity Growth Over Time

Drive Capacity Doubles every two years.
Spinning Disks are getting bigger but not faster!

Requests per GB now exceeds access density capability of spinning drives for many workloads

De-duplication increases access density.

Thin provisioning increases access density.

SSDs are fast for certain workloads but are still expensive.

Interface speeds are getting faster

Virtualization of servers and storage
Other Trends

- Commoditization of hardware
- Block level virtualization
- Blurring of line between file and block level storage
- Many different kinds of “load balancing” solutions
- Server virtualization
- Move to clouds
- Big Data and scalability issues
- Encryption requirements
- Drive capacity increasing faster than error correction rates.
Disk Technology
## Service Time

<table>
<thead>
<tr>
<th>Drive Type</th>
<th>Protocol</th>
<th>Seek</th>
<th>Latency</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATA</td>
<td>1?</td>
<td>9</td>
<td>4.1</td>
<td>14</td>
</tr>
<tr>
<td>10k RPM Fibre</td>
<td>0.3?</td>
<td>4.7</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>15k RPM Fibre</td>
<td>0.2?</td>
<td>3.6</td>
<td>2</td>
<td>5.8</td>
</tr>
<tr>
<td>SSD</td>
<td>0.2?</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
</tr>
</tbody>
</table>

- **Protocol time**
  - Very small < 0.5 ms

- **Average seek, assuming fully used HDD**
  - Range 3.6 – 10 ms depending on technology

- **Latency**
  - Range 2 – 5 ms

- **Data transfer for 512 bytes**
  - Very small

- **Total service time for read**
  - From 0.2 to 15 ms
### Example: Thin Provisioning Capacity

<table>
<thead>
<tr>
<th></th>
<th>Thick Provisioning 300 GB drives</th>
<th>Thin Provisioning 600 GB Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total User Space made available</td>
<td>100 TB</td>
<td>100 TB</td>
</tr>
<tr>
<td>Used space on LUNs</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>Allocated on back-end storage</td>
<td>100 TB (all user space)</td>
<td>40 TB (30% of 100 TB + safety margin)</td>
</tr>
<tr>
<td>RAID Groups required</td>
<td>50 (2TB = 1 RAID Group @ 300 GB)</td>
<td>10 (4 TB = 1 RAID Group @ 600 GB)</td>
</tr>
</tbody>
</table>

**Thick requires 5x the capacity!**

**What about the performance?**
## Thin Provisioning - Performance

<table>
<thead>
<tr>
<th></th>
<th>Thick Provisioning</th>
<th>Thin Provisioning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300 GB drives</td>
<td>600 GB Drives</td>
</tr>
<tr>
<td>Total User I/O ops per second per 100 TB</td>
<td>25,000</td>
<td>25,000</td>
</tr>
<tr>
<td>Available Drives</td>
<td>50 RAID Groups 400 drives</td>
<td>10 RAID Groups 80 drives</td>
</tr>
<tr>
<td>I/Os per drive per second</td>
<td>25,000 / 400 = 62.5 ops/sec/drive</td>
<td>25,000 / 80 = 312.5 ops/sec/drive</td>
</tr>
<tr>
<td>Performance</td>
<td>Good</td>
<td>Will not work..</td>
</tr>
</tbody>
</table>

Access Density for Thin = 5x Thick

Will Thin work for your workload?

Storage Performance Management
© 2012 Storage Networking Industry Association. All Rights Reserved.
Extent Level Automated Tiering

Is the analysis period looking at the right time frame?

What happens to dynamic workloads?

Step 1: Collect I/O Statistics Data
- Tier 0: SSD
  - Vol1: Extent 1
  - Vol2: Extent 1
  - Vol3: Extent 1
  - Vol4: Extent 1
  - Vol5: Extent 1
  - Vol6: Extent 1

Step 2: Analyze I/O Statistics
- Volume | Extent | Read Misses
  - Vol1: Extent 1 | 3 | 45802
  - Vol2: Extent 1 | 6 | 32891
  - Vol3: Extent 1 | 5 | 2702
  - Vol4: Extent 1 | 4 | 5666
  - Vol5: Extent 1 | 2 | 5283
  - Vol6: Extent 1 | 3 | 572

Step 3: Create Migration Plan
- Volume | Extent | Tier
  - Vol1: Extent 1 | 2 | 3
  - Vol2: Extent 1 | 3 | 6
  - Vol3: Extent 1 | 1 | 5
  - Vol4: Extent 1 | 1 | 4
  - Vol5: Extent 1 | 1 | 6
  - Vol6: Extent 1 | 0 | 1

Step 4: Migrate Extents
- Tier 0: SSD
  - Vol1: Extent 2
  - Vol2: Extent 2
  - Vol3: Extent 1
  - Vol4: Extent 2
  - Vol5: Extent 1
  - Vol6: Extent 2

- Tier 1: SATA
  - Vol1: Extent 1
  - Vol2: Extent 1
  - Vol3: Extent 2
  - Vol4: Extent 2
  - Vol5: Extent 1
  - Vol6: Extent 2

Storage Performance Management
© 2012 Storage Networking Industry Association. All Rights Reserved.
RAID
RAID Considerations

- RAID 1 = Mirrored
- RAID 1+0 = Mirrored and striped
- RAID 5 = Single parity
- RAID 6 = Double parity
- Distributed RAID = Extremely wide distribution of parity and data.
- RAID5/RAID6 designed for rebuild times in minutes
- How long does it take to rebuild a single drive failure?
- How long would it take to rebuild a double drive failure?
RAID Rebuild Samples

Calculating RAID5 parity rebuild for SATA
- Estimate = Size of Drive/(I/O Rate * Transfer Size)

Sample:
- Drive Size = 2 TB = 2,000 GB = 2,000,000 Mbytes
- I/O Rate = 400 I/Os per second (Assume Sequential Write with Write Back)
- Transfer Size = 256 Kbytes = .256 Mbytes
- Rebuild Time = 2,000,000/(400*.256) = 5.4 Hours!
- What happens if another failure occurs during rebuild?
- Assumes drive is fully populated
- Assumes 256 Kbytes is the size of the average I/O on disk.
- Assumes no additional new I/O accepted during rebuild.
Storage System Architectures
All vendors agree:

- Frontend ports (host) and
- Backend ports
- Physical (disks) and
- Cache and
- Volumes are required
Processors and Cache

- Different implementations use different approaches
- All use cache to store
  - Recently used tracks and records
  - Recently written records
  - Pre-loaded tracks for sequential read
  - Some form of track descriptor tables to facilitate write operations without a disk access
  - Async copy information
- How effective is cache in relationship to workload?
  - Read cache hit ratio
  - Write delays
We want to know how utilized the Front end adapters are.

- Utilization = I/O Rate * Service Time
- But we don’t always have Service Time!
- Vendor specific clues – Write delays, Queue Length, Slot Collisions
- Cache hits

Front-end Ports may include:

- I/Os and Kbytes/sec
- Response time
- We usually have the throughput capacity, but it can be overstated.
- Estimate Bandwidth Utilization; (Kbytes/sec)/.8 * Rated Capacity
Device Adapters

- Connect HDDs to internal disk system resources
- Manage RAID operations, sometimes using cache memory for RAID computations
- Configured in pairs to provide redundancy if one adapter fails
- HDD interfaces include various generations of SCSI, SSA, FC-AL, SATA and SSD
- FC-AL switched back-end are gradually being replaced by SAS back-ends
- We want to know how utilized the backend disk adapter is?
- Not all vendors provide backend adapter statistics:
  - Utilization = I/O Rate * Service Time
  - But we don’t always have Service Time!
  - We have throughput capacity and we have throughput
  - Estimate bandwidth utilization = KBytesTransferred/Capacity*.8
Disk Drive

- Spinning drives limited on IOPS, operations in milliseconds
  - Mechanically constrained
- SSD drives operations in microseconds
  - Interface and capacity constrained
- Measurements may include:
  - Queue depth, service time
  - Read/Write IOPS
  - Read/Write throughput
  - Utilization
Logical Components: Storage Pools

- RAID Groups are created from Physical Disks
- Storage Pools created from RAID Groups
- Volumes created from Storage Pools
- Typically throughput, response times, I/O rates available for front-end and back-end
Logical Volumes

Volume 1

RAID 5

Disk Drives

Disk 1
- Parity 123
- Data 4
- Data 7
- Data a
- Parity def

Disk 2
- Data 1
- Parity 456
- Data 8
- Data b
- Data d

Disk 3
- Data 2
- Parity 789
- Data 5
- Data c
- Data e

Disk 4
- Data 3
- Data 6
- Data 9
- Parity abc
- Data e

Volume 2

RAID 5

Disk Drives

Disk 5
- Parity 123
- Data 4
- Data 7
- Data a
- Parity def

Disk 6
- Data 1
- Parity 456
- Data 8
- Data b
- Data d

Disk 7
- Data 2
- Parity 789
- Data 5
- Data c
- Data e

Disk 8
- Data 3
- Data 6
- Data 9
- Parity abc
- Data e
Volume Measurements

- Read/write transfer rates, read/write data rates, read/write transfer sizes, read/write cache hit rates, and sometimes read/write response times.
Data Collection
## SPM Technology: Vendor Specific Versus Vendor Neutral

<table>
<thead>
<tr>
<th>Feature</th>
<th>Vendor Specific</th>
<th>SMI-S (Vendor Neutral)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication between tool and storage device</td>
<td>Vendor proprietary fast protocol</td>
<td>Standard XML based communication for all exchanges</td>
</tr>
<tr>
<td>Support for platform specific metrics</td>
<td>No distinction between standard and platform specific metrics</td>
<td>Requires minimum number of fields, but supports vendor extensions for extra fields and components</td>
</tr>
<tr>
<td>Compatibility with other hardware</td>
<td>Specific to one hardware platform</td>
<td>Any hardware supporting SMI-S</td>
</tr>
<tr>
<td>Third party access</td>
<td>Few documented interfaces, different interfaces for each platform</td>
<td>Open specification to any tool vendor</td>
</tr>
<tr>
<td>Provisioning Support</td>
<td>Designed for specific hardware</td>
<td>Difficult to support with SMI-S due to hardware specific implementations</td>
</tr>
</tbody>
</table>
SMI-S is a vendor independent protocol to manage storage subsystems via a HTTP/HTTPS based protocol.

- Common standard across all vendors
- Provide both topology and performance information
- Standard scope includes storage, switches, and servers
- Performance measurement is commonly supported

http://www.snia.org/tech_activities/standards/curr_standards/smi
SMI-S Implementation

Vendor choice

SMI-S Provider

SMI-S (on TCP/IP)

Software exploiting SMI-S

SMI-S Client
Storage system performance measurements need to include:

- Logical Write
- Logical Read
- Physical Write
- Physical Read
- Cache
- Front-end Adapter
- Back-end Adapter
- Disks
# Statistics Summary by ElementType

<table>
<thead>
<tr>
<th>Statistic Property</th>
<th>Top Level Computer System</th>
<th>Component Computer System (Front-end)</th>
<th>Component Computer System (Peer)</th>
<th>Component Computer System (Back-end)</th>
<th>Front-end Port</th>
<th>Back-end Port</th>
<th>Volume (Logical Disk)</th>
<th>Composite Extent</th>
<th>Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic Time</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>TotalIOs</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>Kbytes Transferred</td>
<td>R</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>IOTimeCounter</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>ReadIOs</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>ReadHitIOs</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>ReadIOTimeCounter</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>ReadHitIOTimeCounter</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Kbytes Read</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>N</td>
</tr>
<tr>
<td>Write IOs</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>WriteHitIOs</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>WriteIOTimeCounter</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>N</td>
</tr>
<tr>
<td>WriteHitIOTimeCounter</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>KbytesWritten</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>IdleTimeCounter</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>MaintOp</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>MaintTimeCounter</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**R** - Required  
**O** - Optional  
**N** – Not Specified
Relationships – A simple example

- Configuration data is necessary to provide the relationships between the elements:
  - Which LUNs are defined on which extent pool
  - Which physical drives make up an array group
  - Which port (types) are connected to each (host adapter)
Summary

- The goal of storage performance management (SPM) is to reduce storage costs while maintaining performance SLAs.

- SPM consists of:
  - Processes
  - Measurement
  - Skills

- SMI-S provides a solid foundation for obtaining the necessary measurements to implement SPM
Please send any questions or comments on this presentation to SNIA: tracktutorials@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

- Brett Allison
- Gilbert Houtekamer
Bridging the Visibility Gap

An effective Storage Performance Management (SPM) solution must:

- Provide visibility inside the storage system where 70% of the bottlenecks occur, and
- Automatically correlate your workload metrics with the specific hardware component capabilities.