Storage Performance Management Overview

Brett Allison, IntelliMagic, Inc.
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]
SPM is the process of ensuring
- that users constantly receive required I/O service levels to avoid performance problems;
- that storage assets are efficiently configured and used to avoid over spending on hardware.

Risk avoidance is more important than saving money.
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

Lower

→ size (and cost) of storage configuration

Cost

Performance

Utilization levels

Lower
The Four SPM Primary Processes

- **Watch Health**: Early Warning, Historical Trending
- **Find Constraint**: Root Cause, Imbalance
- **Buy Hardware**: Upgrade, Replace
- **Move Data**: Within Hardware, To New Hardware
## SPM Maturity Stages

<table>
<thead>
<tr>
<th>SPM Maturity Stage</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>
SMI-S
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
### What ElementTypes Does SMI Define for Performance?

<table>
<thead>
<tr>
<th>ElementType</th>
<th>Component</th>
<th>Vendor Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElementType2</td>
<td>Cumulative statistics for the storage system</td>
<td></td>
</tr>
<tr>
<td>ElementType3</td>
<td>Front-end Controllers</td>
<td></td>
</tr>
<tr>
<td>ElementType4</td>
<td>Peer Storage System (Mirroring)</td>
<td>✔</td>
</tr>
<tr>
<td>ElementType5</td>
<td>Back-end Controllers</td>
<td></td>
</tr>
<tr>
<td>ElementType6</td>
<td>Front-end FC ports</td>
<td></td>
</tr>
<tr>
<td>ElementType7</td>
<td>Back-end Ports</td>
<td></td>
</tr>
<tr>
<td>ElementType8</td>
<td>Volumes</td>
<td></td>
</tr>
<tr>
<td>ElementType9</td>
<td>Extent – Intermediate storage</td>
<td>✔</td>
</tr>
<tr>
<td>ElementType10</td>
<td>Disk Drive</td>
<td></td>
</tr>
<tr>
<td>ElementType11</td>
<td>Arbitrary Logical Units – Controller commands</td>
<td>✔</td>
</tr>
<tr>
<td>ElementType12</td>
<td>Remote Replica Group – Remote Mirror</td>
<td>✔</td>
</tr>
</tbody>
</table>
## 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>Total I/Os</td>
<td>R</td>
<td>R</td>
<td>O</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>O</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>Kbytes Transferred</td>
<td>R</td>
<td>O</td>
<td>O</td>
<td>R</td>
<td>O</td>
<td>R</td>
<td>O</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>Read I/Os</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>Read Hit I/Os</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Read IOTimeCounter</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>Read Hit IOTimeCounter</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>O</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>Write I/Os</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>Write Hit I/Os</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Write IOTimeCounter</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>Write Hit IOTimeCounter</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 Written</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>N</td>
</tr>
<tr>
<td>Idle Time Counter</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>Maint Op</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>
<tr>
<td>Maint Time Counter</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
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)
Storage System Architectures and Measurement with SMI-S
Spinning disks are getting bigger faster than they are getting faster!

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

Trends driving density increase include:
- De-duplication
- Thin provisioning

SSD could remove disk as bottleneck in the future.

Commodity hardware
All vendors agree:
- Frontend (host) and
- Backend (disk) and
- Cache and
- Volumes are required

Do the metrics provided tell us what we need to know?
- SMI-S has a well defined model for performance metrics and relationships.

High level storage system metrics are not enough!
- We can only rely on these if the I/O is evenly spread across components
Throughput and Response time
for all data
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
Read Cache Hit % by DSS

Read hit (%)
for all Disk Storage Systems by Serial
Front-end Adapter

- Provides connectivity between disk subsystem and hosts
- Cards support ESCON, SCSI, FICON Fibre, SAS and/or iSCSI sometimes FICON and Fibre with one card
- Implementations differ greatly in maximum data handling capability, especially for FICON and Fibre
- Even though ports are rated as (e.g.) 8 Gbit/s, no implementation achieves this speed due to overhead.
Front-end Controller/Adapter Read Response Time

Read Response (ms) by Serial

Storage Performance Management
© 2011 Storage Networking Industry Association. All Rights Reserved.
Read and Write Throughput (MB/s)
for all Ports by Serial and Link ID

Front-end Ports
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
Backend Adapter Throughput

Total MB/s for operations handled by Back-End Adapter
by Serial

MB/s

0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200

© 2011 Storage Networking Industry Association. All Rights Reserved.
Logical Components: RAID Group Sets/Storage Pools

- RAID Groups are created from Physical Disks
- RAID Group Sets created from RAID Groups
- Volumes created from RAID Group Sets
Interpreting Disk Response Service Levels

HDD Response Service Levels
For Serial is ' ' by RAID Group Set ID

- Greater than 20 ms
- Greater than 10 ms
- Lower than 10 ms
Volume Performance Metrics

Highest I/O Rate Volumes (top 20)
For RAID Group Set ID is 'extpool36', for Serial is 'DS8100-2' by Volume Label

I/Os per sec

0 2 4 6 8 10 12 14 16 18 20 22 24

AIOB AIOC AICD AICE AJBO AJBI AJAI AJAI ALBD ALBE ALAC AJAC AJAC
Volume – Fast Write Bypass %
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.
Q&A / Feedback

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 Hautekamer
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.