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

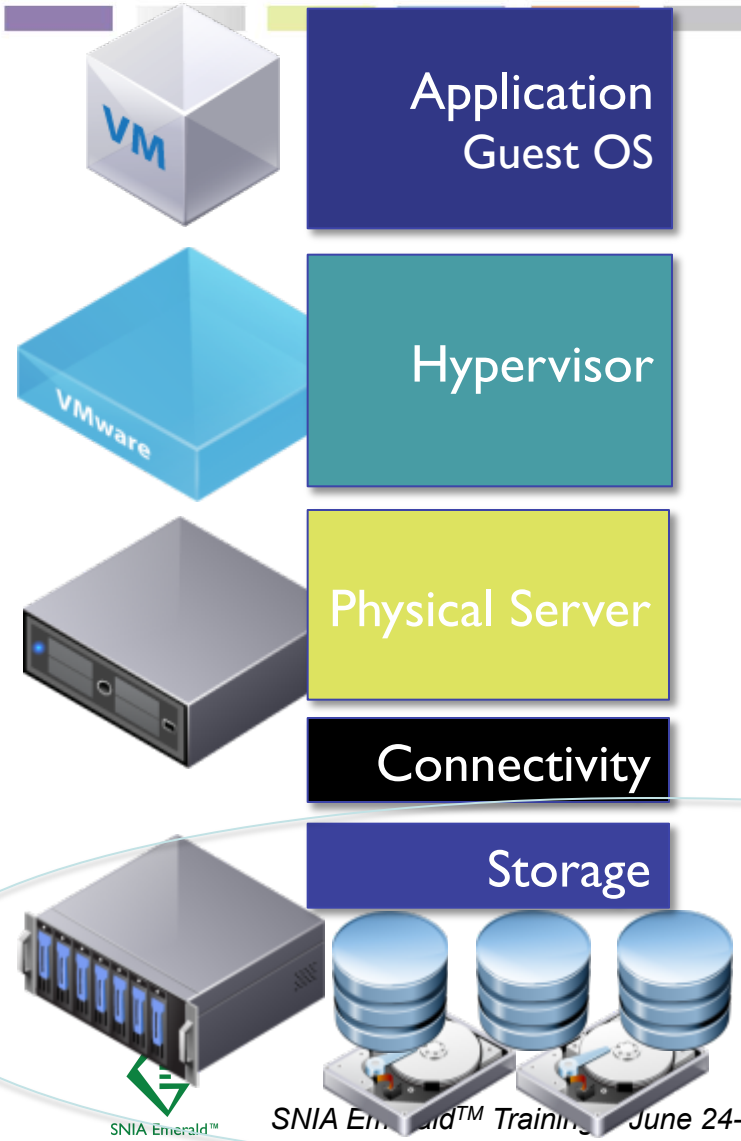
SNIA™ Green Storage Initiative

SNIA Emerald™

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

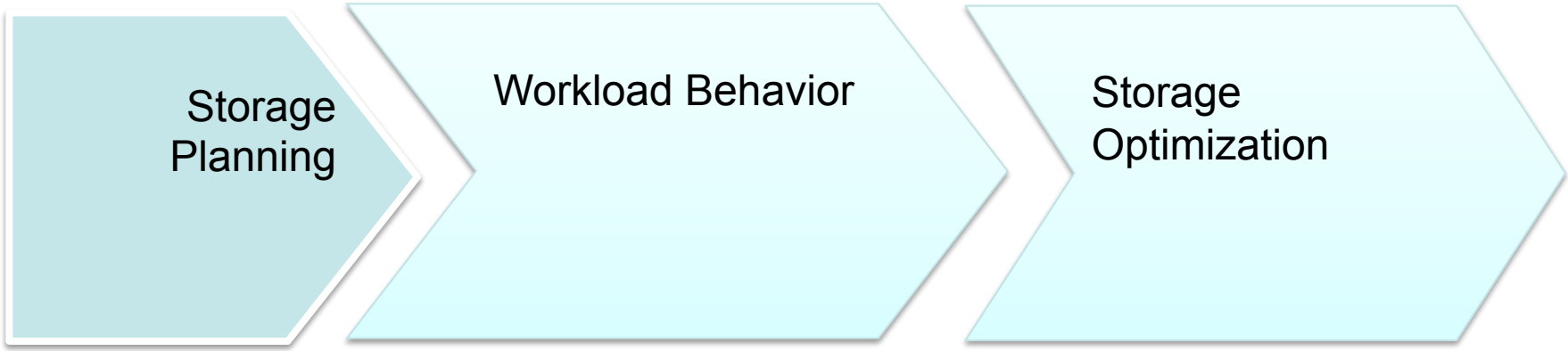
Physical Server Level
CPU and Memory Saturation, Power Saving

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

Storage Level
SAN or NAS Devices
Utilization, Latency, Throughput

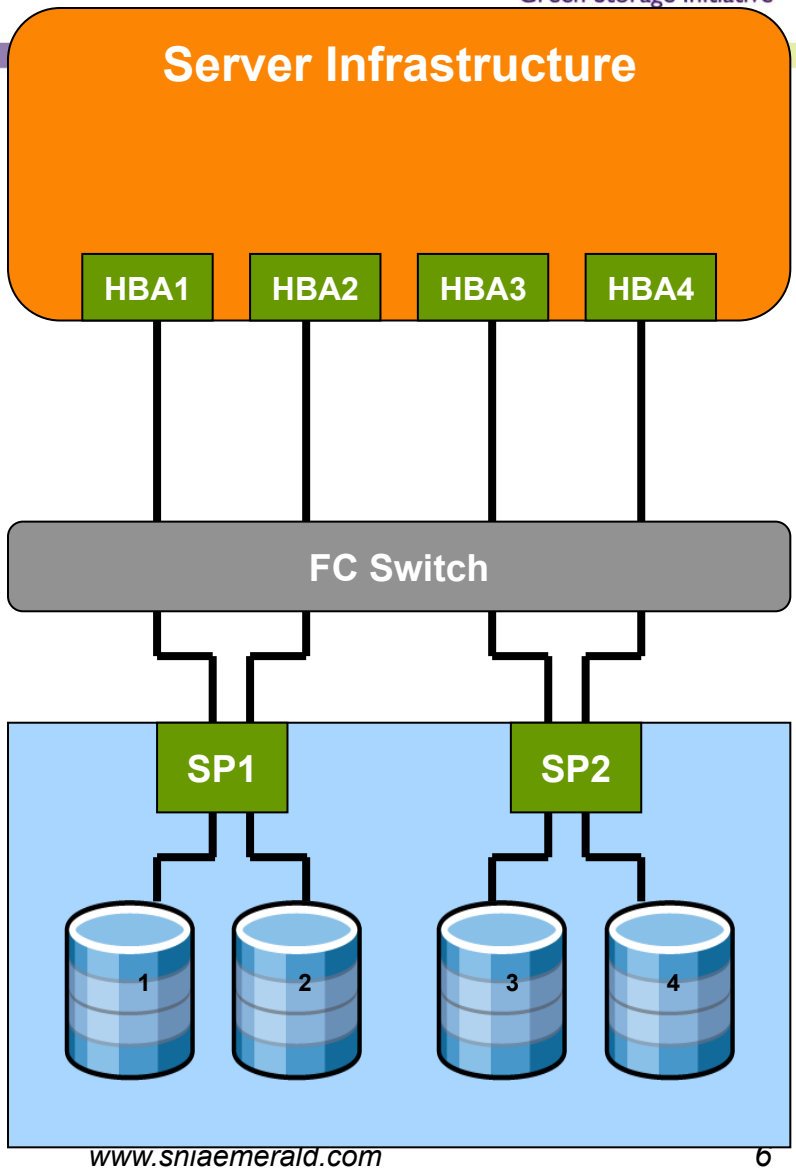
Storage Performance Planning

Planning for Performance



Storage Planning

- 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

Drive Type	Speed	MB/sec	IOPS	Latency	LC Manage
FC 4Gb	15k	150	200	5.5ms	High Perf. Trans
FC 4Gb	10k	75	165	6.8ms	High Perf. Trans
SAS (6Gb,12Gb)	10k	150	185	12.7ms	Streaming
SATA (6Gb,12Gb)	7200	140	38	12.7ms	Streaming/Nearline
SATA	7200	68	38	12.7ms	Nearline

Solid State Storage

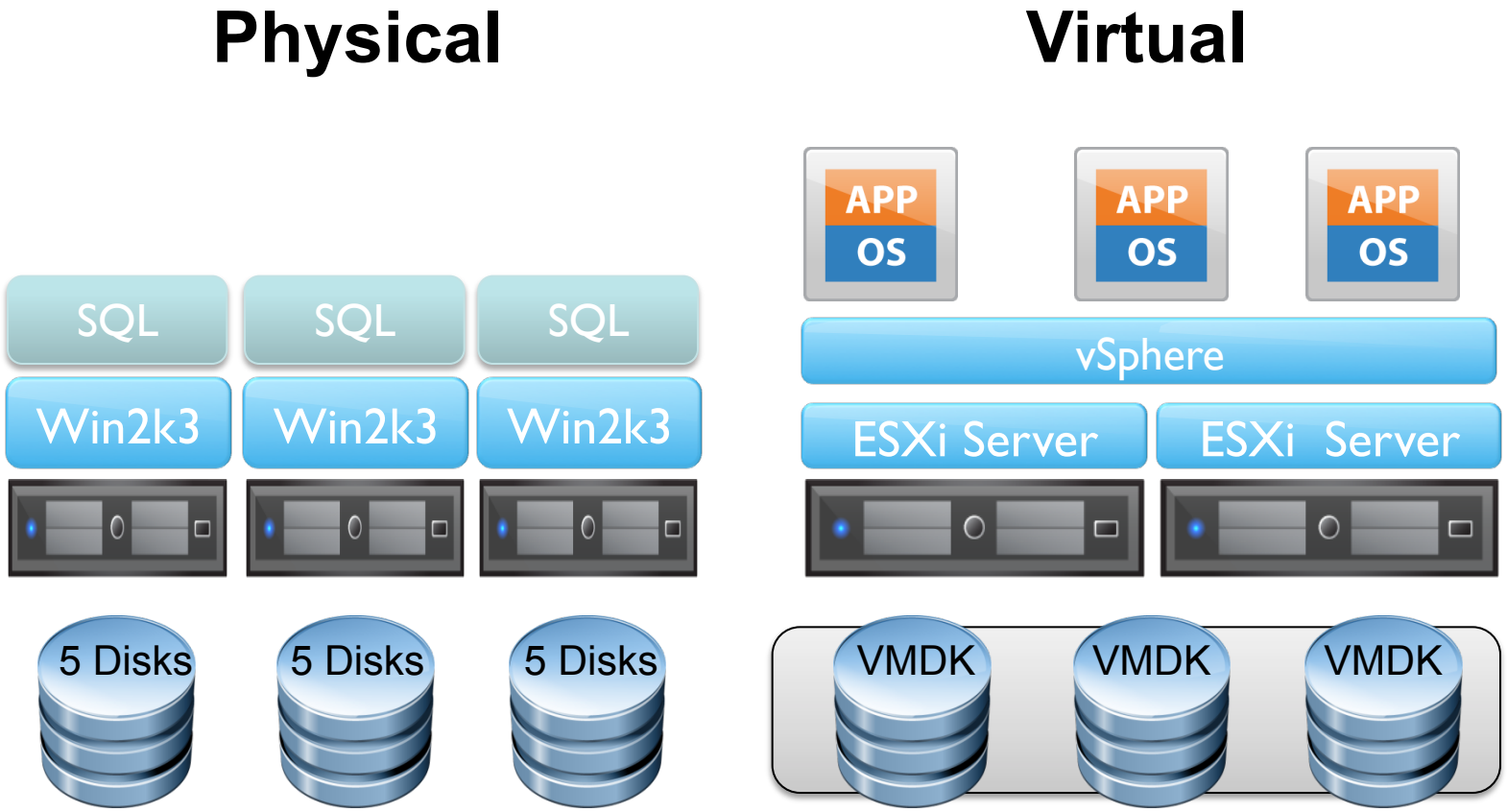
- No all SSDs designed the same
  - NAND-based flash memory
  - DRAM-based (Random Access Memory)
  - Enterprise flash drives (EFDs)
  - Hybrid Drives
- Performance varies widely
  - Capacity
  - Compression
  - Wear leveling
  - Error Correction and bad block mapping
  - Metadata management
  - Garbage collection
  - Encryption

Solid State Storage

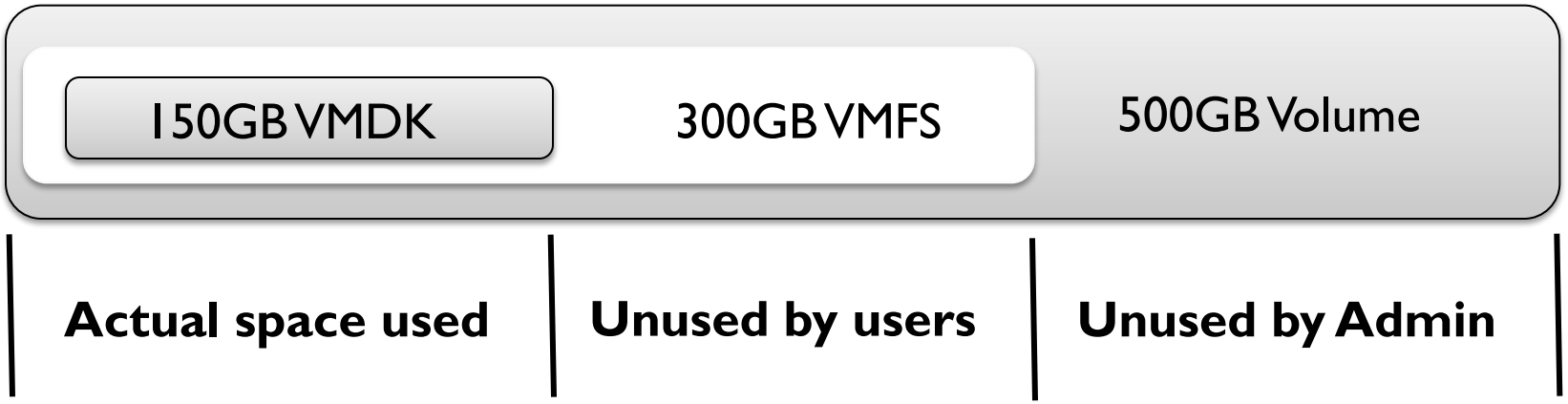
Metric	NAND Flash	
	SLC	MLC
Latency (microseconds)	100	200-300
Persistence	10x more persistent	Less reliable*
Cost	30% more expensive	More cost effective
Sequential read/writes	3x faster	Slower

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

Virtualize to consolidate

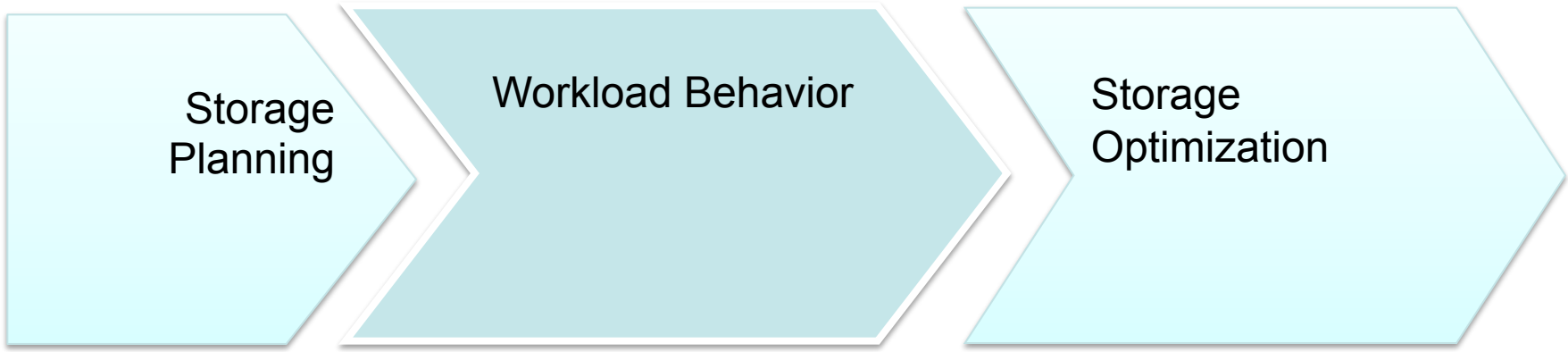


Over Provisioning

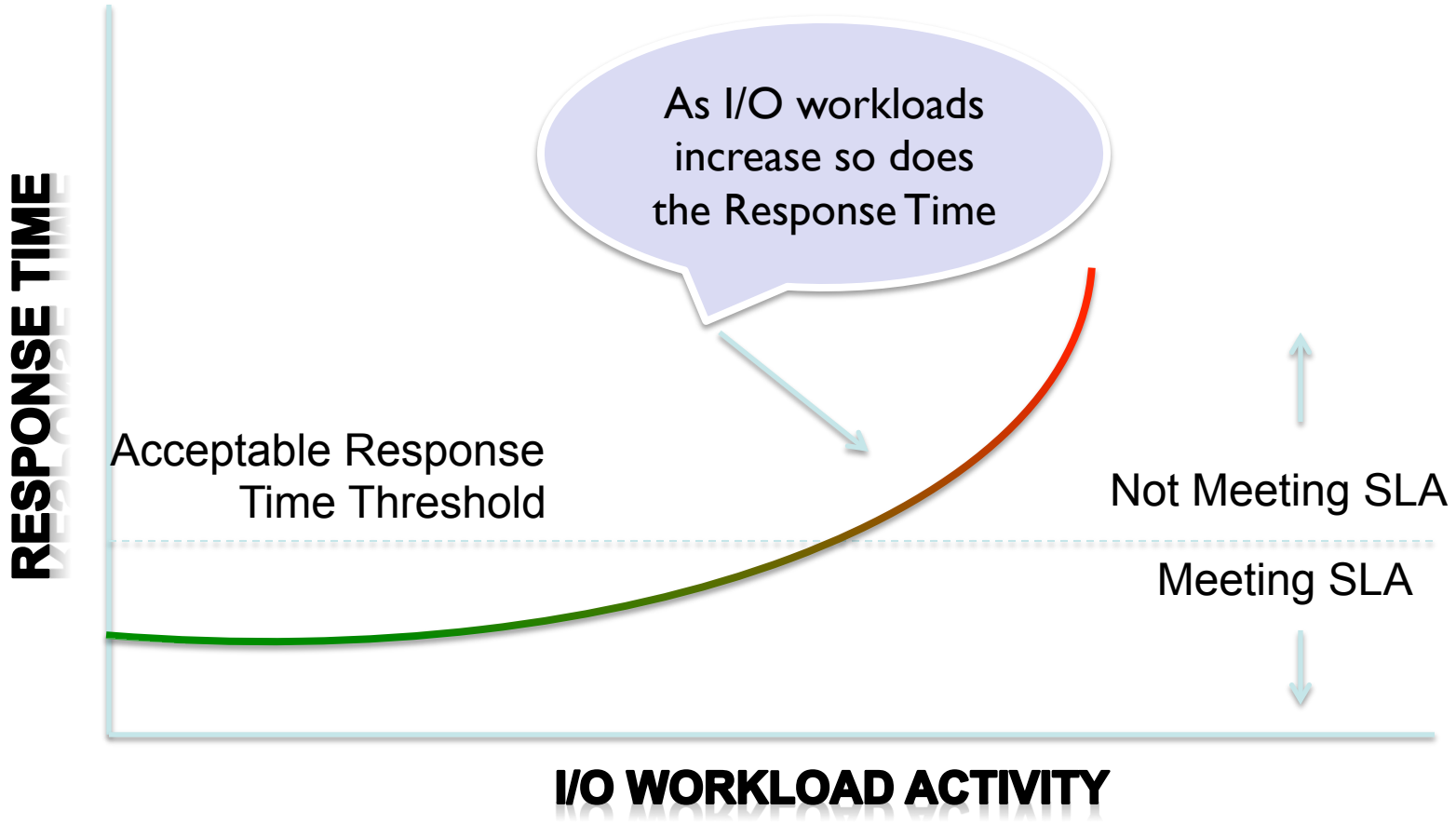


- Using Thick provisioning it is easy to over provision.
- You may want to **consider Thin Provisioning**.
- Most vendors offer Thin Provisioning

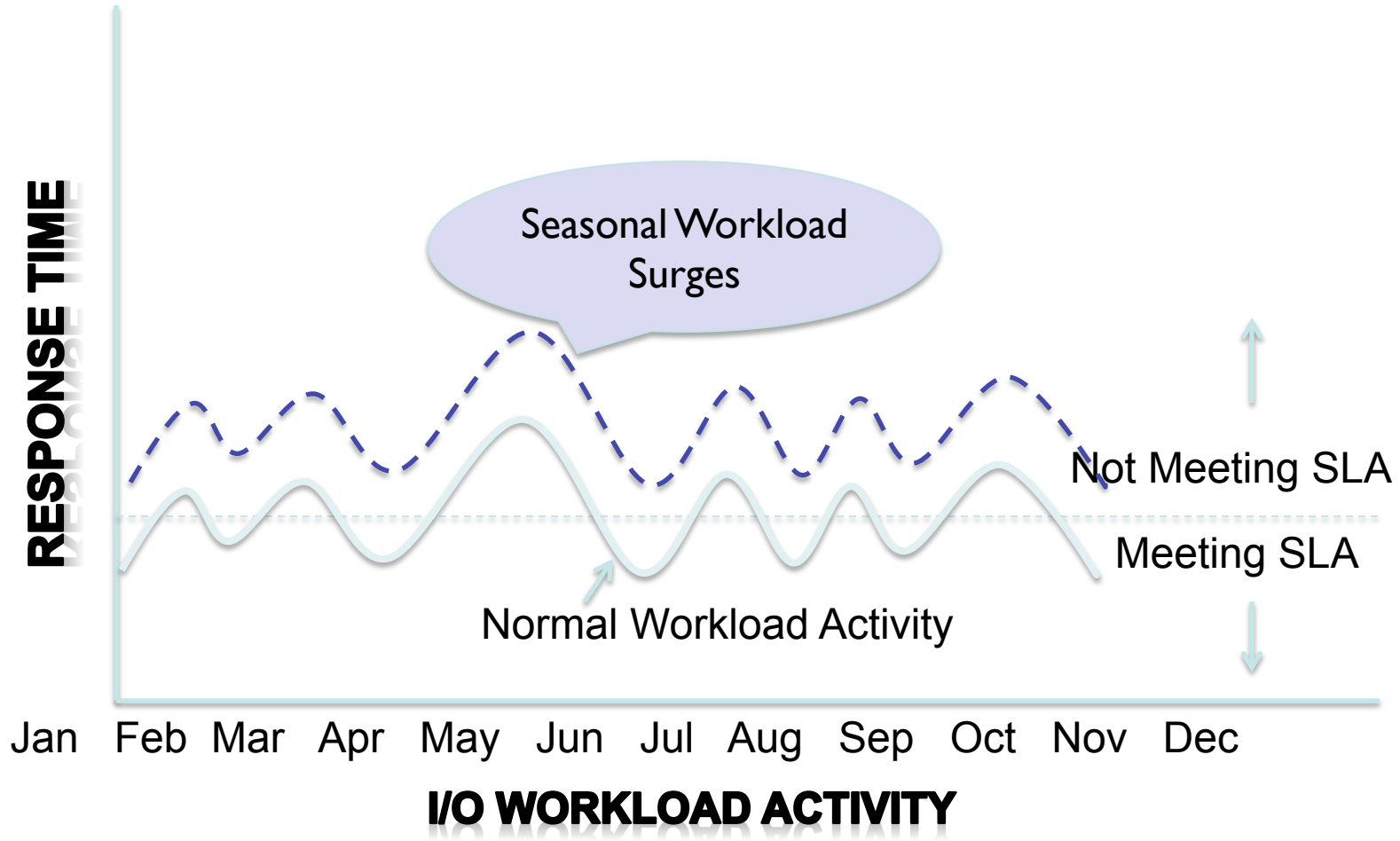
Planning for Performance



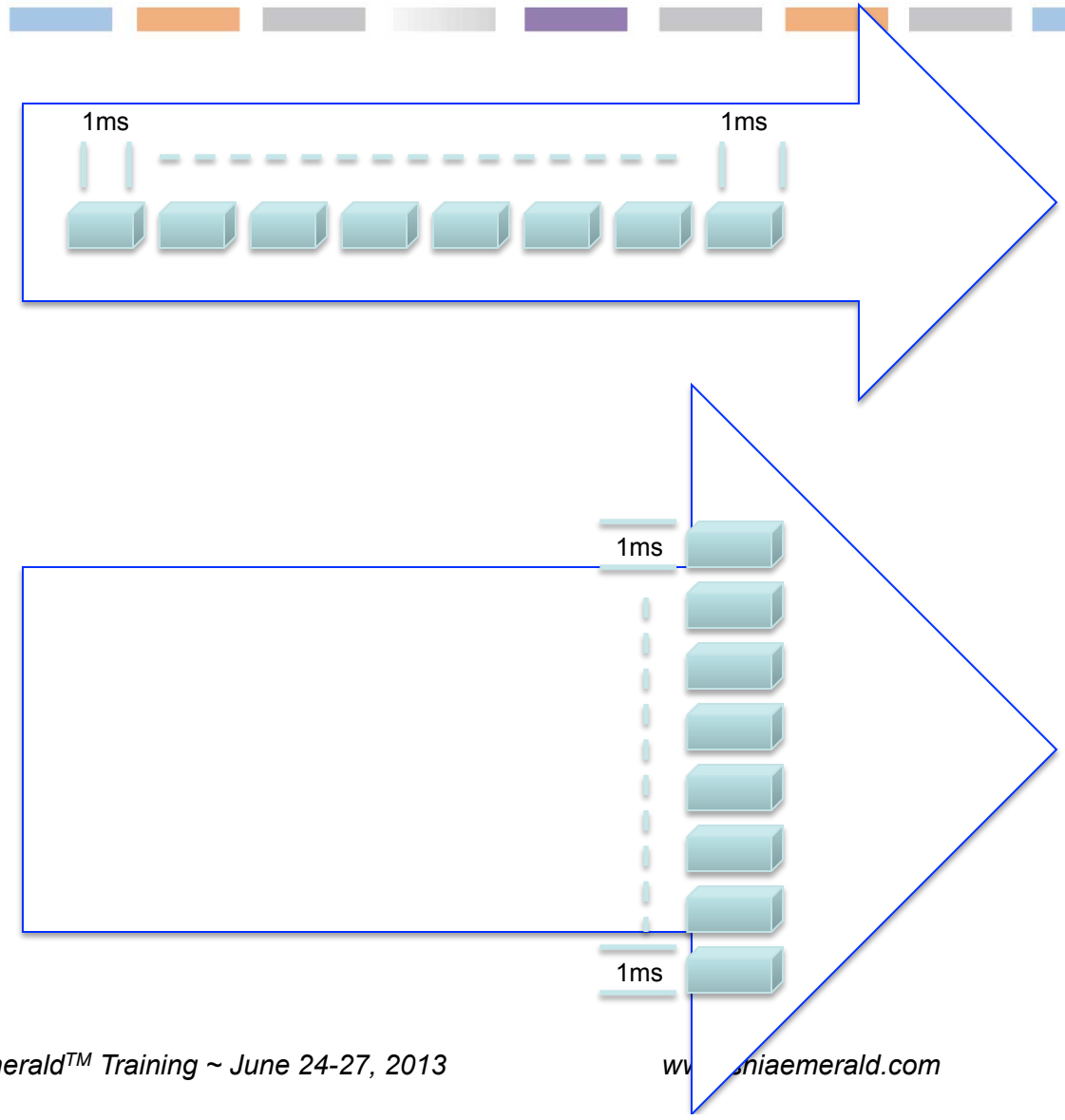
I/O Workload Activity vs. Response Time Supply and Demand



Seasonal/Periodic Performance Surges



Single vs. Multi-threaded Applications



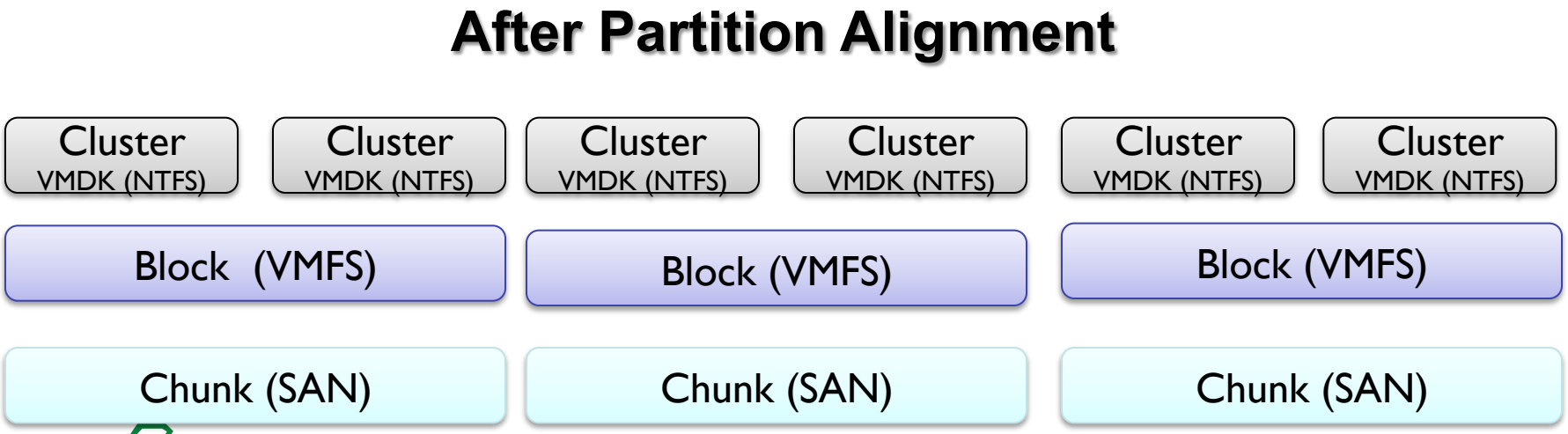
I/O Queue Depth

- 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

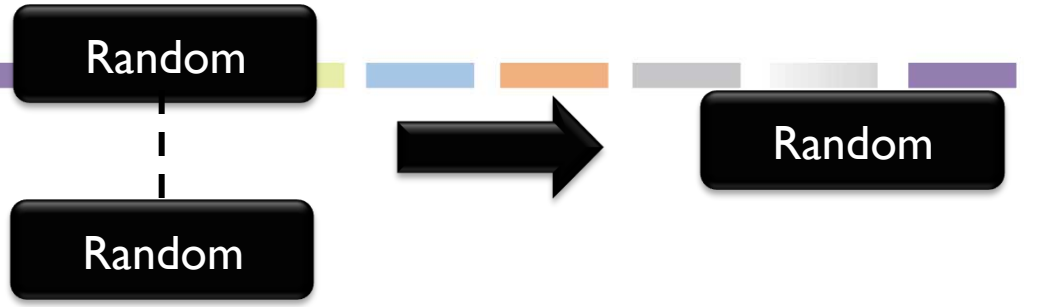
Skew

- 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.

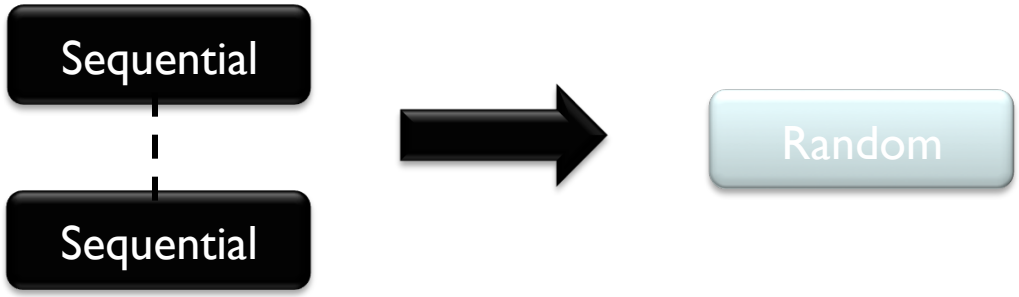
Misalignment



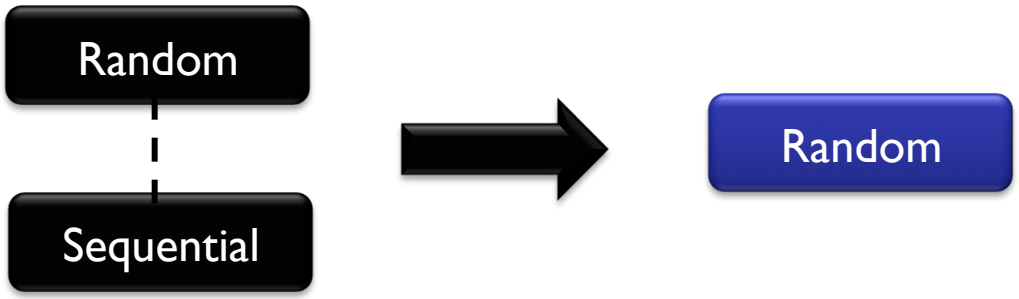
Workload Consolidation



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

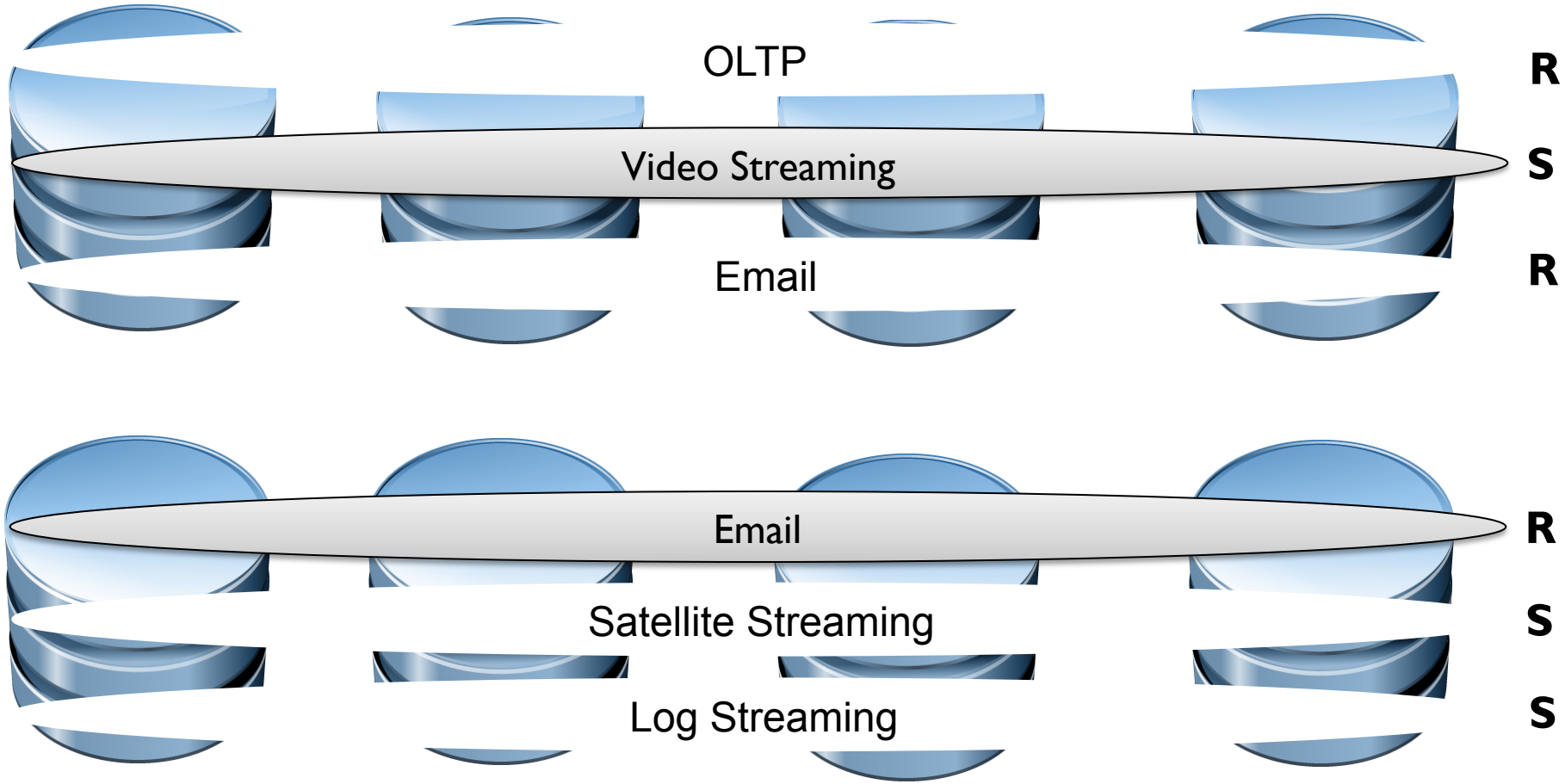


Too many sequential threads on a lun will appear as a random workload to the storage ***Negative Impact on Sequential Perf***.

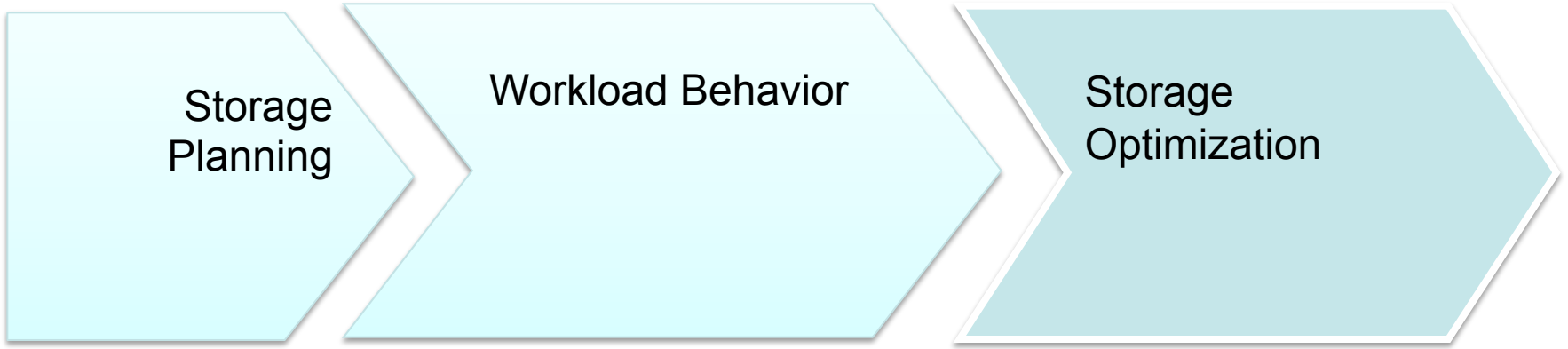


Mixing Sequential with Random can hurt Sequential workload Throughput. ***Negative Impact on Sequential Perf***.

Mixed Workloads

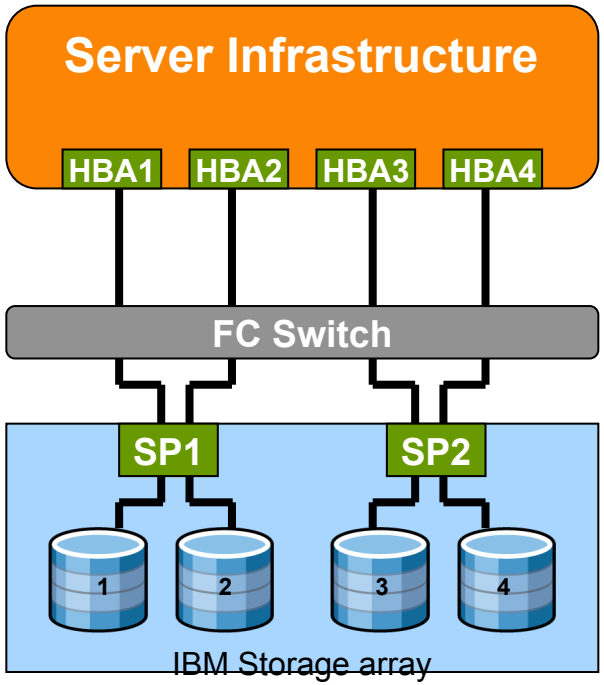


Planning and Best Practices



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



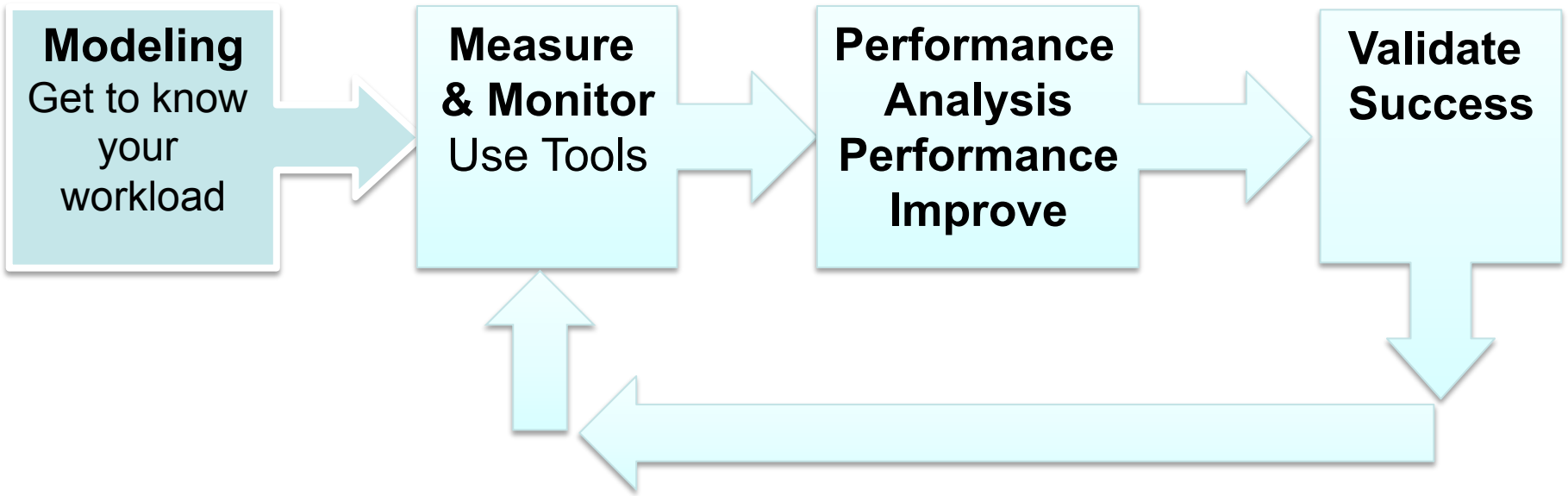
Optimizing Storage

- 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

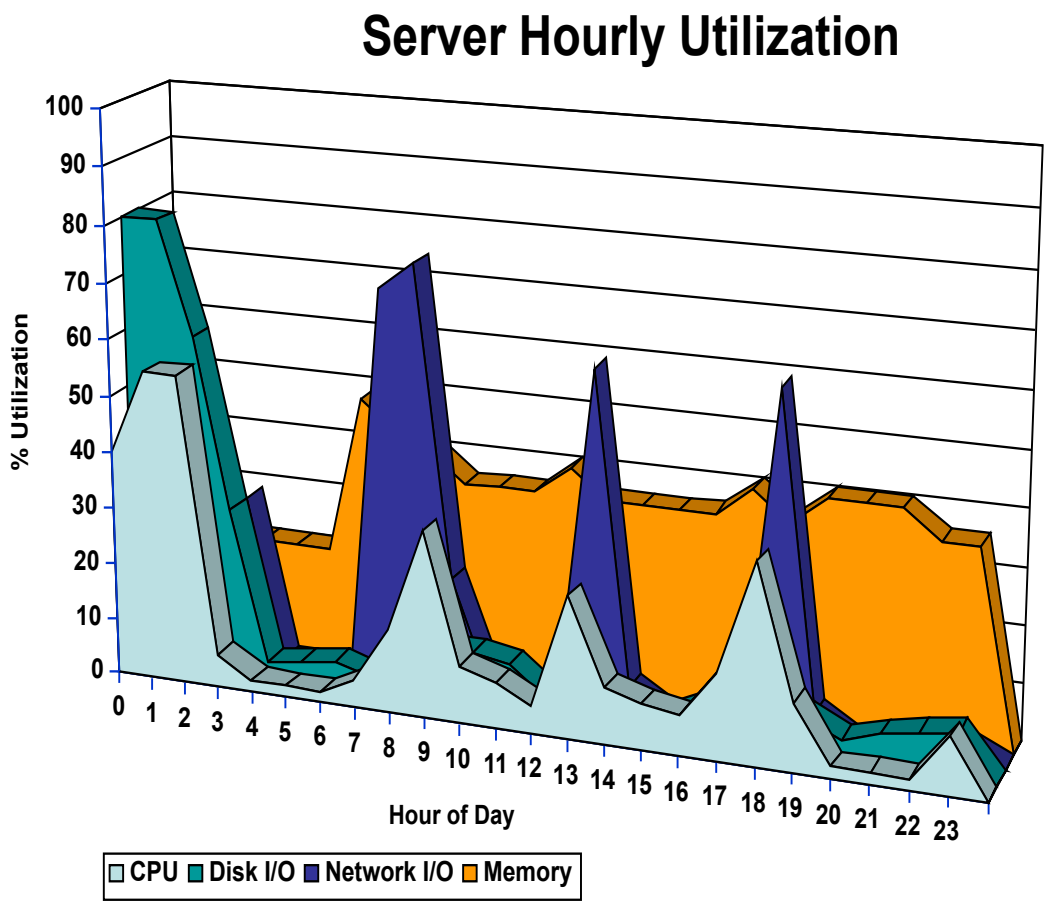
Performance Methodology



Understanding Your Workload

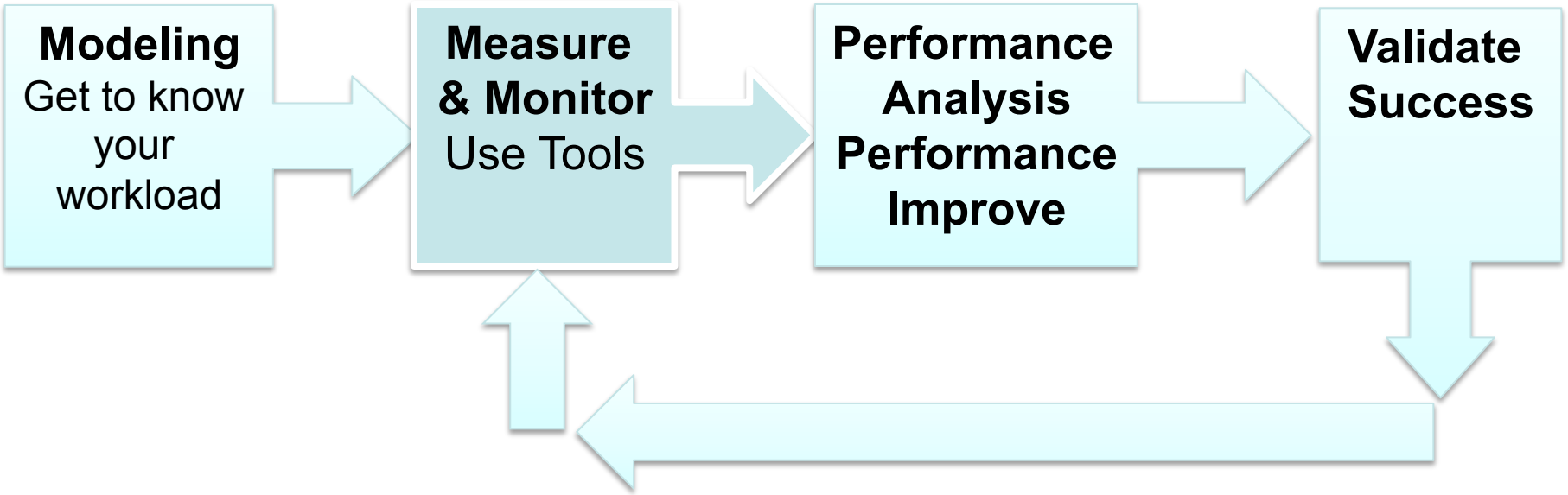
- Workload Indicators
  - Demand for resources vs. Resources currently used
  - Result is a percentage of Workload
    - Low latency number is Good – Object has the resources it needs
    - Can go above 100% - Object is “Starving”
- Workload summarized across critical resources
- Workload Details View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



Analyze all resource dimensions

Performance Methodology



Approach to Real-Time Performance Management

SNIA Green Storage Initiative

3rd Generation – Holistic, Real Time Analytics

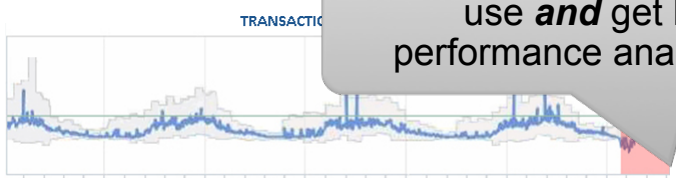
Flexible INTEGRATION to many data sources

gomez · Microsoft System Center Operations Manager · keynote · Tivoli software · EMC smarts · HYPERIC · vmware · ...

Enterprise SCALABILITY

Patented performance ANALYTICS

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



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

Powerful information DASHBOARDS




Infrastructure vs. Operations Impacts on the storage performance & efficiency

OPERATIONS

Performance
App = host
Limited movement

Capacity
Dedicated resources
Pre-committed

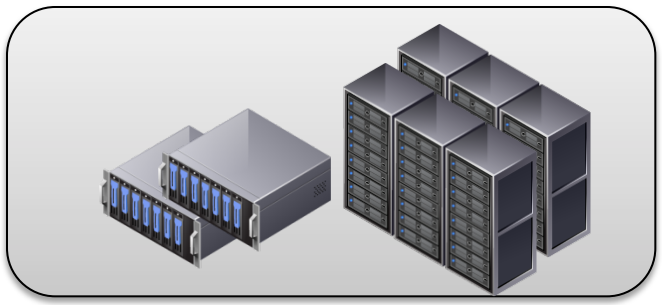
Configuration
Static , pre-configured
1-1 mapping

Performance
Resource abstraction
Mobility
+

Capacity
Resource pooling
Minimize waste, Prevent Stress
+

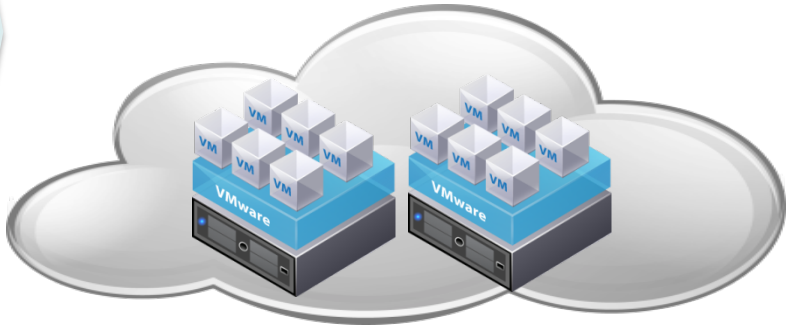
Configuration
On-demand, self-service
Dependencies

INFRASTRUCTURE



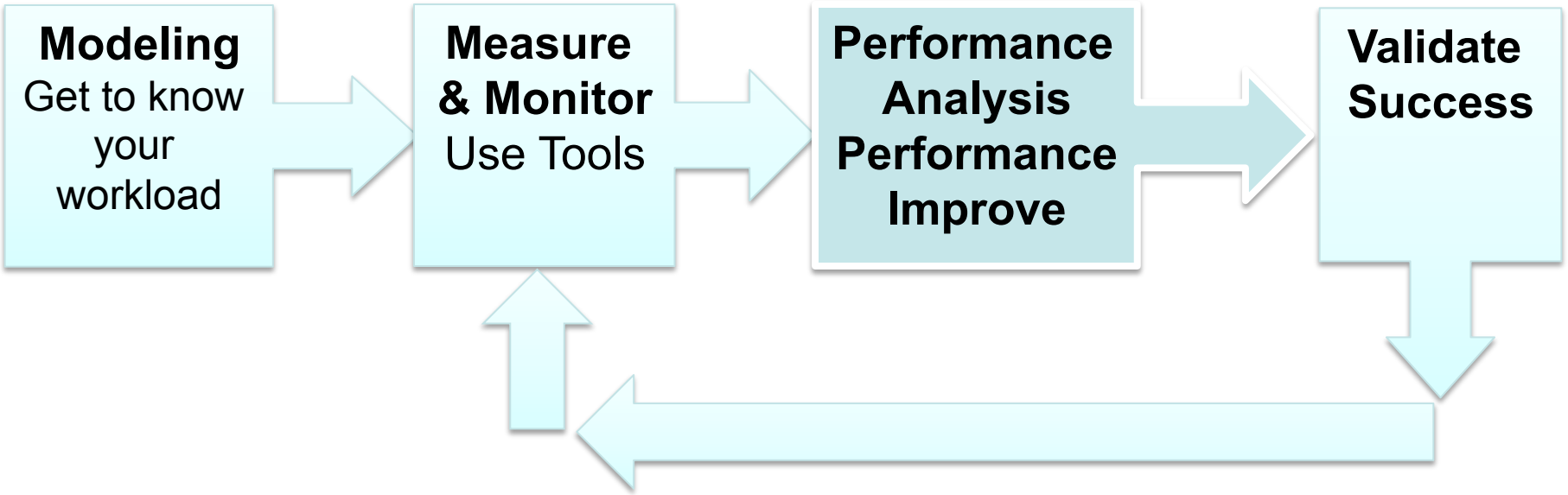
Physical Datacenters

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



Virtual Datacenters

Performance Methodology



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

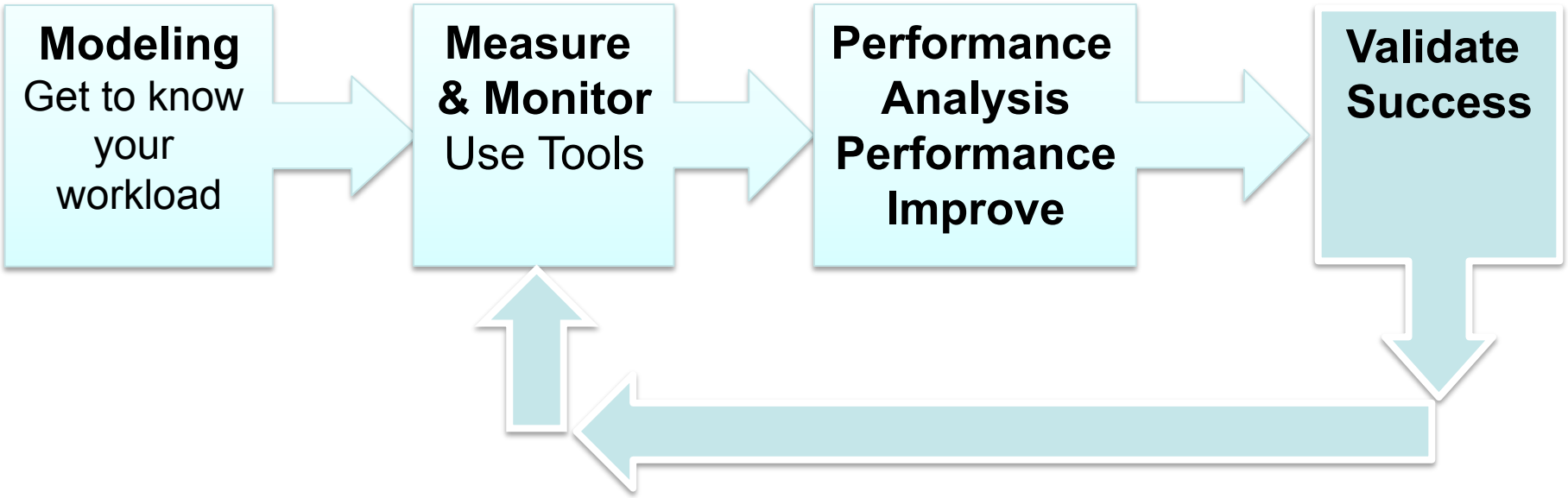
Basic Metrics

- 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

Metric	Described	Threshold
Average Device latency	latencies from the storage system	10-15 ms
Average Kernel latency	Latencies from the kernel's I/O subsystem	1-2 ms
Aborts and retries	Can't keep up with demand and times out or something broke	1
Response Time	Overall application or OS response time	Many IOs above 10 ms

Performance Methodology



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@evaluatorgroup.com
Twitter:@vLeahSchoeb