



Education

What's Old Is New Again – Storage Tiering

Larry Freeman, NetApp, Inc.

Author: Larry Freeman, NetApp, Inc.

- 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.

Summary: The SNIA defines tiered storage as “storage that is physically partitioned into multiple distinct classes based on price, performance or other attributes.”

Although physical tiering of storage has been a common practice for decades, new interest in automated tiering has arisen due to increased availability of techniques that automatically promote “hot” data to high performance storage tiers – and demote “stale” data to low-cost tiers.

Topics will include:

- Fundamentals of Storage Tiering
- Levels of granularity in tiering
- Achieving optimal placement of data.
- Recent innovations in logical and virtual tiering techniques
- Using the cloud as a tier

After viewing this session, attendees should gain understanding in:

- Tiering fundamentals and benefits
- Trends in automated tiering
- Tiering best practices
- Tiering resources

- The SNIA Data Protection and Capacity Optimization (DPCO) Committee was created to foster the growth and success of the market for data protection and capacity optimization technologies.



<http://www.snia.org/forums/dpco>



Check out **SNIA DPCO Tutorials:**

- **Trends in Data Protection and Restoration Technologies**
- **Advanced Data Reduction Concepts**

“It’s almost impossible to conduct a conversation with a storage vendor without raising the topic of “storage tiering.” ... While there is tremendous user value to be gained from employing storage tiering, there is also at times a lack of clarity and of semantic accuracy in the market.”

**Mark Peters,
Enterprise Strategy Group
July 12, 2011**

What is a Storage Tier?

- ◆ A storage media class:
 - ◆ High Performance SSD/Cache
 - ◆ High Performance HDD
 - ◆ High Capacity HDD
 - ◆ Tape
- ◆ A data class:
 - ◆ Mission critical data
 - ◆ Hot data
 - ◆ Cold data
- ◆ A pricing class:
 - ◆ Networked Storage
 - ◆ DAS
 - ◆ Public cloud

What is a Storage Tier?

- A storage media class:
 - ◆ High Performance SSD/Cache
 - ◆ High Performance HDD
 - ◆ High Capacity HDD
 - ◆ Tape
- A data class:
 - ◆ Mission critical data
 - ◆ Hot data
 - ◆ Cold data
- A pricing class:
 - ◆ Networked Storage
 - ◆ DAS
 - ◆ Public cloud

From the SNIA 2011 Dictionary:

tiered storage
[Storage System]

Storage that is physically partitioned into multiple distinct classes based on price, performance, or other attributes.

Data may be dynamically moved among classes within a tiered storage implementation based on access activity or other considerations.

Why Is Tiering Important?

- With so many types of storage devices available today, with varying price/performance characteristics, it is difficult to know which tier should be purchased for a given workload
- Data is dynamic, sometimes needed in an instant while other times not needed at all
- Tiering helps assure that data is always available and accessible at the correct performance level

Storage Choices

| | | |
|-----------|---|------------|
| Very Fast |  | \$\$\$\$\$ |
| Fast |  | \$\$\$\$ |
| Medium |  | \$\$\$ |
| Slow |  | \$\$ |
| Very Slow |  | \$ |

Tiering Considerations

- Users would like their data on the highest tier - for the highest performance
- Businesses would prefer to keep data on the lowest tier – for the lowest cost
- Placing all data in the middle tiers is an unacceptable compromise
- Tiering is an attempt to solve this problem – by placing data in the right tier at the right time to satisfy both cost and performance concerns

Storage Tiers

| | | |
|--------------|---|------------|
| Flash |  | \$\$\$\$\$ |
| FC/SAS |  | \$\$\$\$ |
| SATA |  | \$\$\$ |
| Tape |  | \$\$ |
| Public Cloud |  | \$ |

Tiering Choices

Traditional Tiers

- Tier 1 – 15K FC / SAS Disk Drives
- Tier 2 – 7.5K ATA/SATA Drives
- Tier 3 – Tape Drives

Emerging Tiers

- Tier 0 – SSD Drives
- Storage Cache as a Tier
- Public Cloud as a Tier
- Server Cache as a Tier

Tiering Choices

Traditional Tiers

- Tier 1 – 15K FC / SAS Disk Drives
- Tier 2 – 7.5K ATA/SATA Drives
- Tier 3 – Tape Drives

Emerging Tiers

- Tier 0 – SSD Drives
- Storage Cache as a Tier
- Public Cloud as a Tier
- Server Cache as a Tier

Price/Performance Examples (2011)

| Storage Media | Performance* | Cost/GB** |
|--------------------------------------|--------------|-----------|
| PCI Flash | 1500 MB/sec | \$26.17 |
| SSD | 500 MB/sec | \$1.98 |
| SAS HDD | 200 MB/sec | \$0.69 |
| SATA HDD | 150 MB/sec | \$0.04 |
| LTO-5 Tape Cartridge | 140 MB/sec | \$0.04 |
| Public Cloud | 2-?? MB/sec | varies |
| *Sustained Data Transfer Rate | | |
| ** Published Price of Component Only | | |

Source: SNIA DPCO Committee

Storage Tiering - 1980s

Mainframe DFHSM (Data Facility Hierarchical Storage Manager)

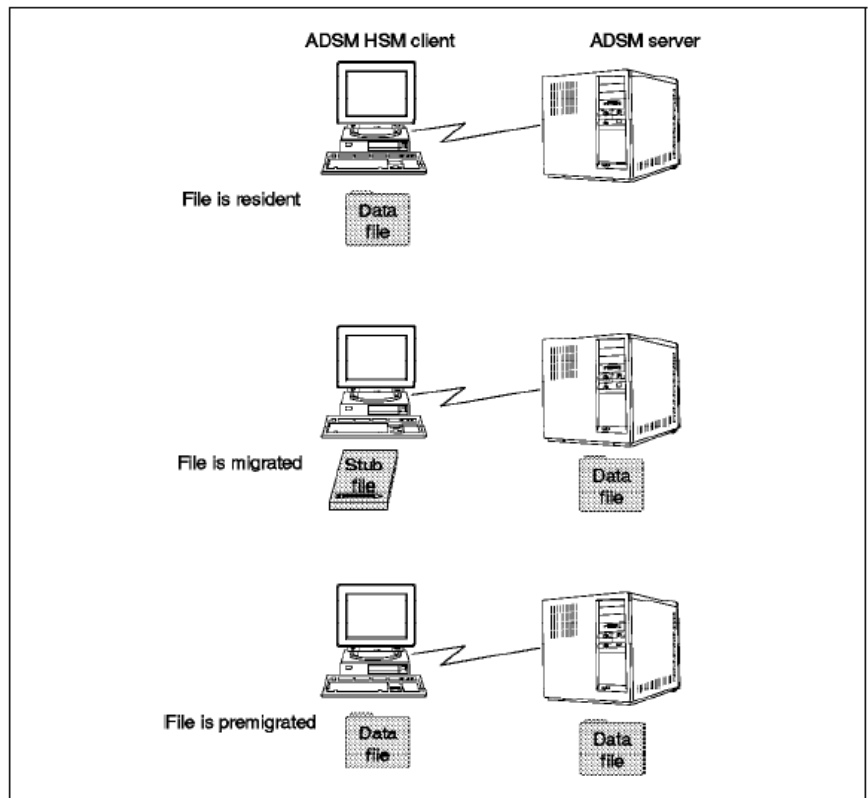
- Born out of necessity
- Online disk was cost prohibitive
- Not enough capacity to store active data
- Automatic staging and retrieval from tape using file descriptors and migration levels
 - Online Disk
 - Compacted Disk (Mig Level 1)
 - Tape (Mig Level 2)
- Files were migrated to MLI AND ML2 based on threshold and age
- High/Low threshold could be set by user
- Often sold with IBM 3850 Mass Storage System or STK Nearline tape library



Storage Tiering – 1990s

HSM (Hierarchical Storage Management)

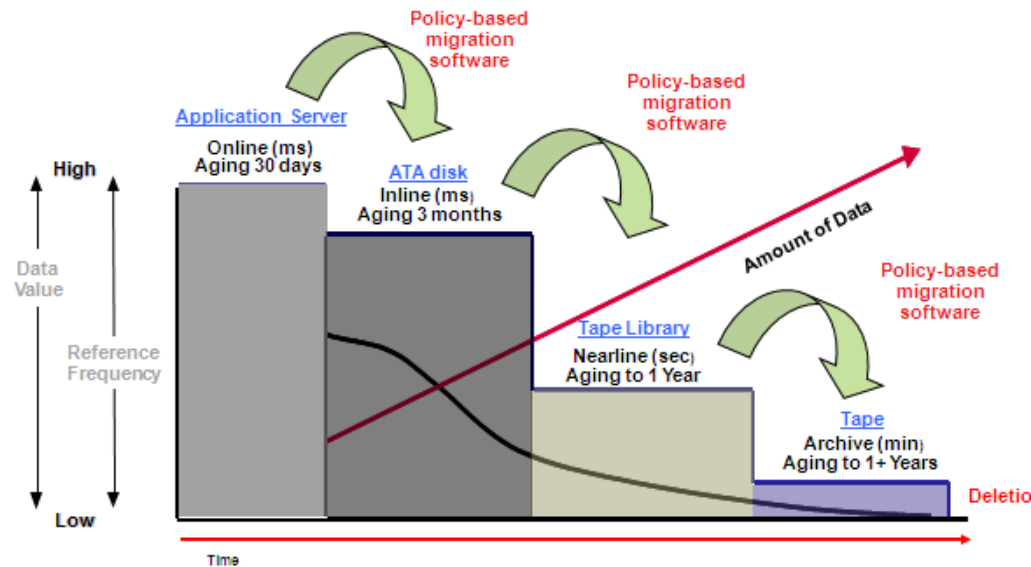
- Eligible file-based data migrated to a dedicated HSM server
- Migrated files could be backed up directly from HSM server
- Migration policy engine:
 - ◆ File Size
 - ◆ File Age
 - ◆ File Type
 - ◆ File Location
 - ◆ Capacity Threshold
 - ◆ Scheduled Migration



Storage Tiering – 2000s

ILM (Information Lifecycle Management)

- Not a single product, but a process
- Best practices for lifetime data retention – including compliance policies
- Minimum requirements:
 - ◆ Policy Engine
 - ◆ Data Mover
 - ◆ Tiered storage hierarchy
- Proposed for both file and sub-file data



Storage Tiering Today

1. Automated Tiering

- File and sub-file migration performed transparently within one storage array or between multiple arrays
- Uses predictive analysis for migration

2. Storage Cache Tiering

- Promotion of “hot” data to cache
- Data concurrently resides in cache and on disk

3. Cloud Tiering

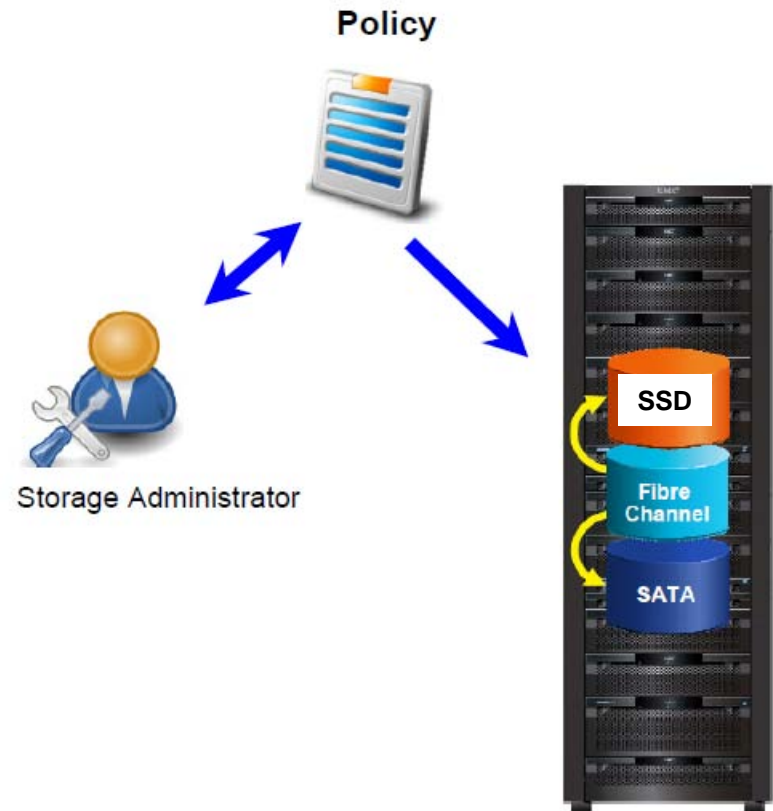
- Using cloud gateways
 - ◆ Software or hardware
- Push inactive data to the cloud
- Use public cloud as a backup tier

4. Server Cache Tiering

- Host-based cache
- Some data intelligence
- Places data closest to application

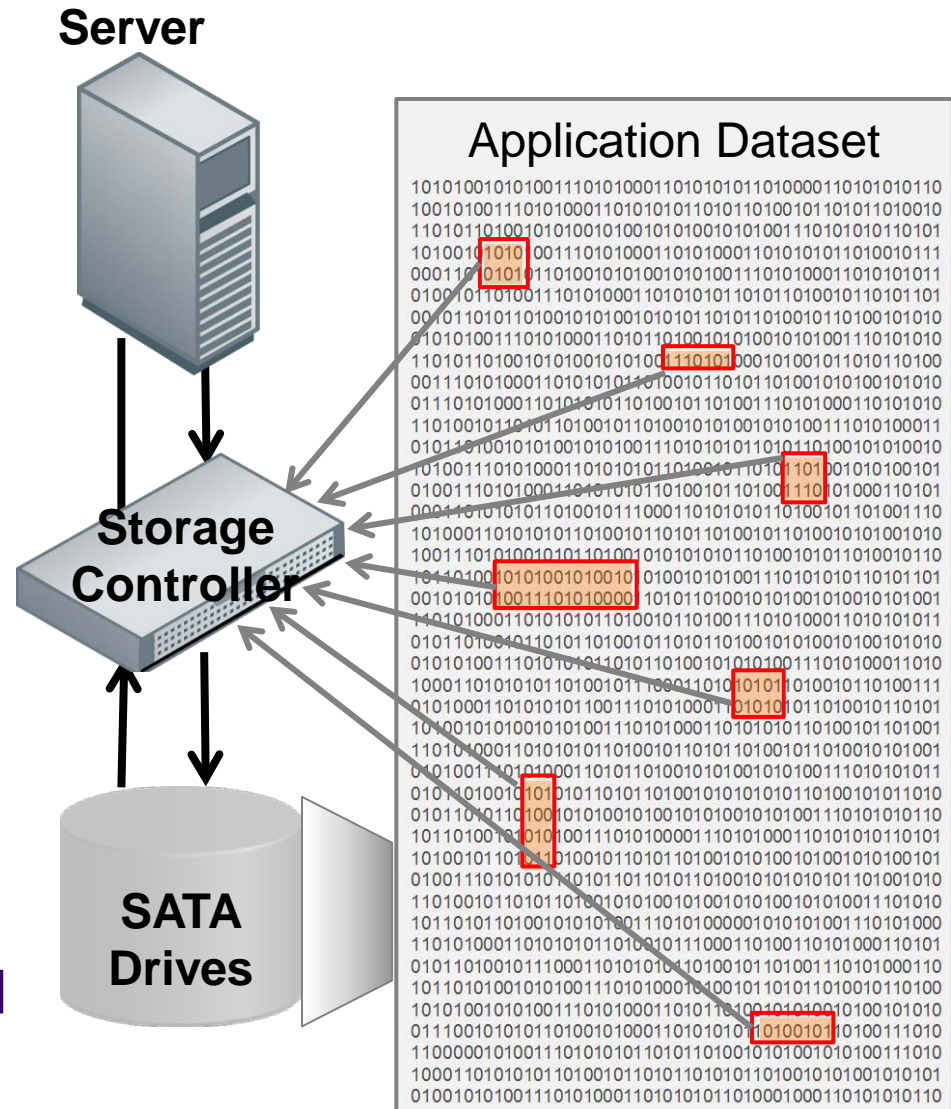
1. Automatic Tiering

- Data is migrated between tiers based on policies and data usage patterns
- Improves value and efficiency of SSD by placing only frequently accessed data on expensive storage media
- Tiering can be on a single system or across multiple systems
- Will this remove the need for high performance HDD's?



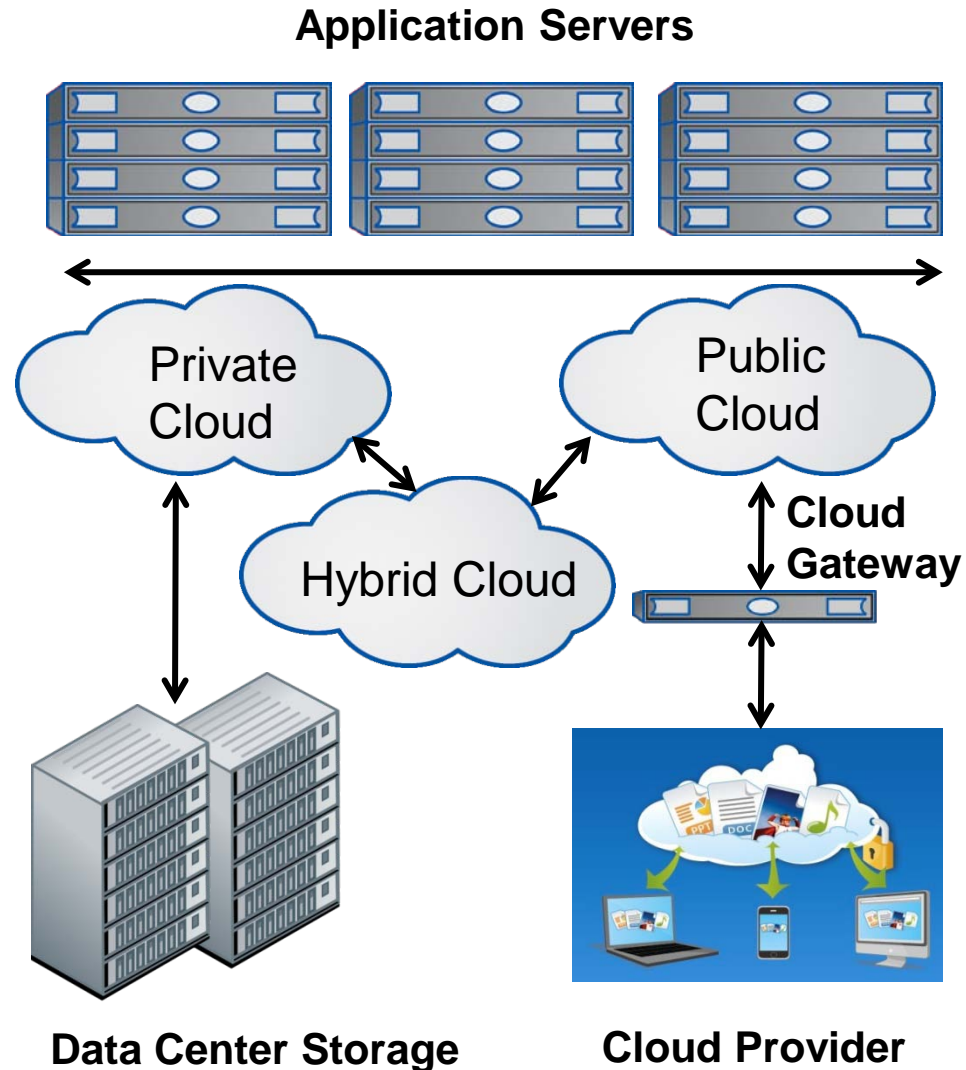
2. Storage Cache Tiering

- Cache used as an alternative to traditional tiering
 - ◆ Sometimes called “virtual” tier
- “Hot” sub-file data automatically promoted to cache
- Original data remains in disk
- Allows capacity HHD’s to be used in place of performance HDD’s
- Reduction in storage costs and spindle count



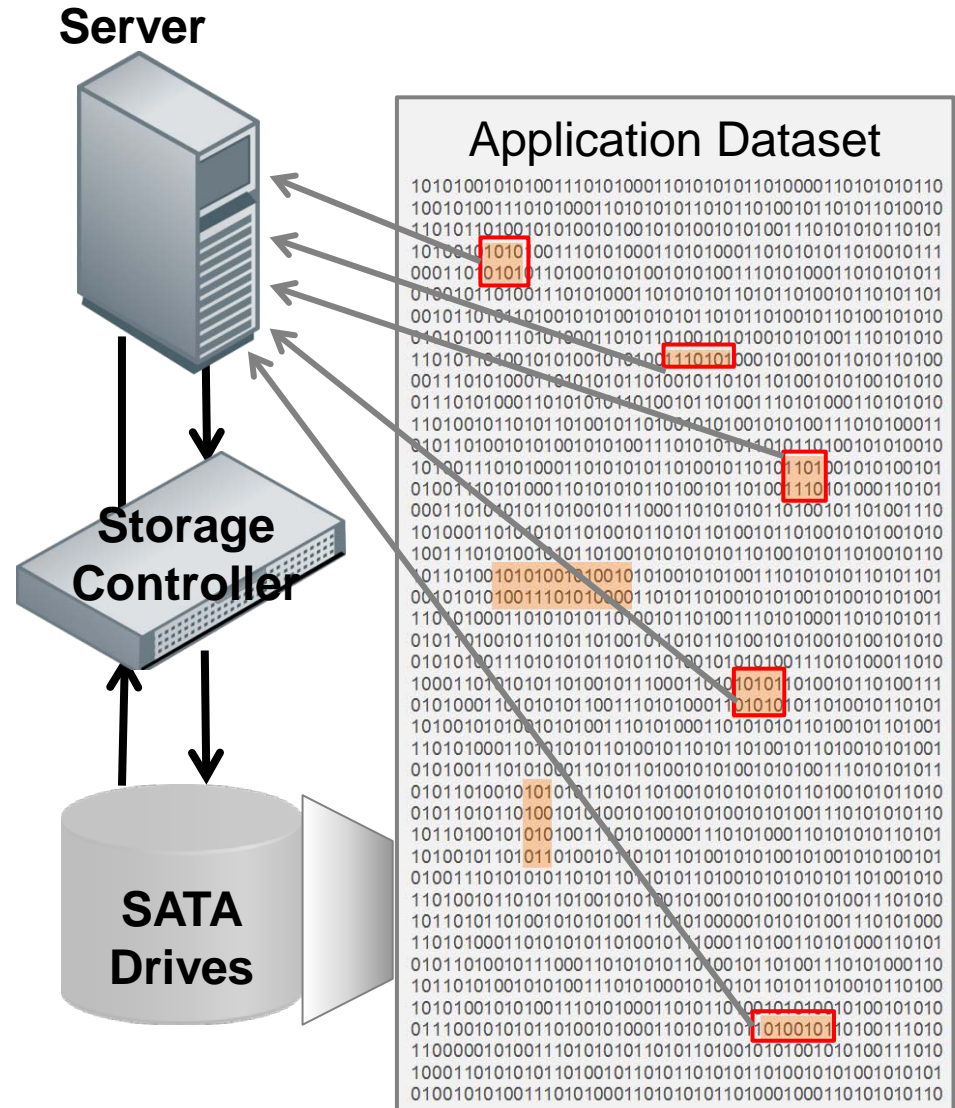
3. Cloud Tiering

- Cloud Gateways are being developed for direct block and file access
- Cloud is being used as an automated backup tier or manual archival tier
- Ultimately a Hybrid Cloud will emerge which will combine public and private clouds using a common data management protocol (CDMI)



4. Server Cache Tiering

- Similar to Storage Cache
- Hot data cached to Server-hosted Flash
- Reduced Latency
- Managed by Storage Controller
- May use write-through or write-back cache with read-through
- Some challenges with efficiencies in cached data
 - ◆ Deduplication
 - ◆ Compression



- Determine the problem you are trying to solve
 - ◆ I want fastest possible throughput
 - ◆ I want lowest possible cost
 - ◆ I want the best combination of the above
- Where do you want to implement tiering?
 - ◆ All of my storage arrays
 - ◆ Some of my storage arrays
 - ◆ Some of my applications
 - ◆ Some of my data
- Understanding your objectives will make your selection easier

Questions To Ask During Selection

➤ Tiering Scale and Granularity

- Is data tiered within an array – or across arrays?
- Is data tiered at the file or sub-file level?

➤ Tiering Policies and Methods

- ◆ Reactive, Predictive, or Realtime?
- ◆ How often is data re-tiered?
- ◆ Can the tiering policy be adjusted?

➤ Tiering Costs

- ◆ Additional software or hardware required?
- ◆ What is the installation process?
 - › Self Install?
 - › Professional Services Required?

- Today's tiering technologies are rooted in prior generations...
 - ◆ Hierarchical Storage Management (HSM)
 - ◆ Information Lifecycle Management (ILM)
- ...but are substantially different

- New storage technologies and storage array intelligent are bringing new capabilities
- Knowing your objectives and which questions to ask will make your job easier

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

- SNIA Education Committee

SNIA Data Protection and Capacity Optimization Committee (DPCO)

Mike Dutch

Larry Freeman

Gene Nagle

Ron Pagani

Michael Peterson

Thomas Rivera

John Tyrrell

Joseph White

Send any questions or comments on this presentation to SNIA: tracktutorials@snia.org