

*The Information Company for Storage Professionals*

**Evaluator Group**

# Storage System Tiering



# Agenda

- Tiering Technology for Selected Vendors

- Flash vs. Cache

- Tiering Implementation & Messaging for Selected Vendors

- Compellent
- Dell EqualLogic
- EMC VNX
- EMC VMAX
- HP 3PAR
- HDS VSP
- IBM (DS8000 & V7000)
- NetApp
- Xitech



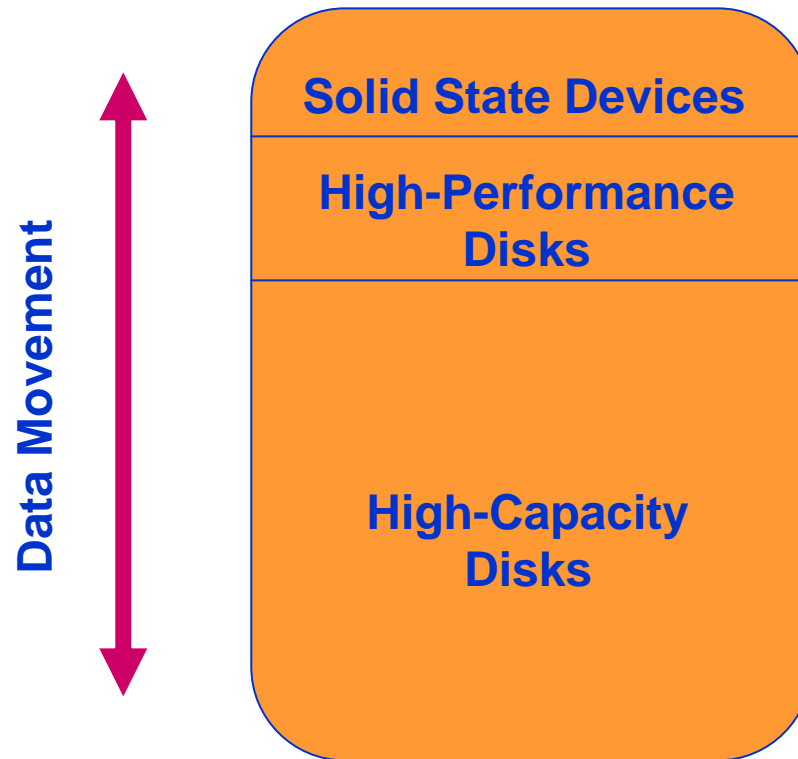
# Tiering vs. Caching

- Storage Tiers
  - Implies a particular price and performance metric
  - Provides actual capacity
  - All content resides on media
  - Performance is limited only to media speed
  - May be improved with Caching
- Caching
  - Not considered actual storage
  - All capacity must be backed by non volatile media
  - Performance limited to size of cache
  - Limited use for Random workloads

# Flash as Cache

- Pros for Flash as Cache
  - Relatively low cost way to add performance
  - Automated, very little tuning required
  - Highly effective for sequential read workloads
  - Somewhat effective for random write workloads
- Cons for Flash as Cache
  - Not as effective for random workloads, particularly reads
  - Adds cost, without adding capacity
  - More workloads are becoming more random
    - VMware drives seemingly random workloads to storage
    - VDI workloads are moderately random

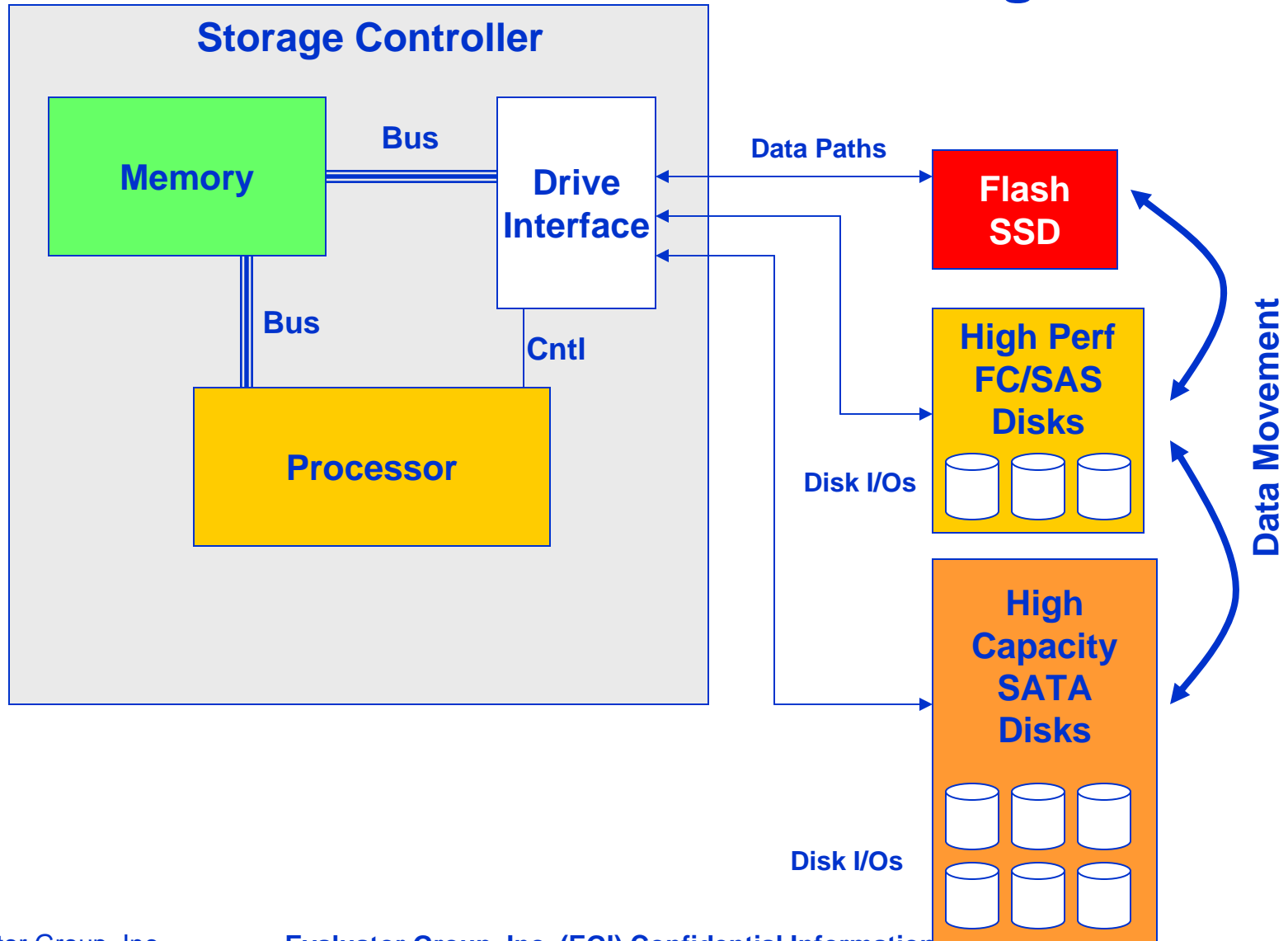
# Storage System Tiering “Within-the-Box”



**Volume or sub-LUN level**

**May include virtualized external storage**

# Tiering Data



# Compellent Data Progression

- Architecture
  - System captures in-depth access information
  - Storage profile for SSD's required
  - Supports 9 tiers of storage
    - Based on 3 drive types, 3 rotation speeds, 3 RAID levels
  - Based on System datapage size
  - Page size can be either 512K, 2 MB (default) or 4 MB
- Data Progression
  - Default is 3 days of learning move to promote data, 12 days for demotion
  - Write cache should be disabled on SSD volumes
  - Reports built-in to system, including amount of data to be moved up or down next cycle
  - Considers RAID levels, and track placement as well

# Compellent Messaging

- Data Progression
  - Part of Fluid Data Storage overall message
- Component of Storage Efficiency top level
  - Thin provisioning
  - Automated storage tiering
  - Continuous snapshots
  - Disk pooling (termed storage virtualization)
- Optimize enterprise storage
  - Automated
  - Eliminate manual data classification
  - Maximize write performance with SSDs

# Dell EqualLogic AutoTiering

- Architecture
  - Starts after 15 minutes – up to 8 hours
  - Includes multiple tiers
  - Volume level for E, X, V and S
  - Sub-volume for XVS models
    - Uses 8 @ 100 GB SSD, + 8 @ 450 GB 15k
    - Supports only 2 tiers
  - Tiering Criteria
    - Drive type
    - RAID Level
    - I/O Activity
    - Random I/O

# Dell EqualLogic AutoTiering

- Organization of EqualLogic System
  - Can create up to 4 storage pools across members (called Arrays which are storage controllers with drives) in a PS Series Group (also called a PS Series SAN)
  - Storage Pools can have
    - Different RAID protection from members (members with different RAID)
    - Different types of drives from members
  - Storage pools can be used for physical separation of data
  - Volumes are created/assigned to a storage pool
    - Volumes can be manually moved between storage pools
    - No load balancing between pools

## **Dell EqualLogic AutoTiering**

- XVS systems support sub-LUN (sub-volume tiering)
  - Two tiers supported – SSD and 15K SAS drives
  - Built on using EqualLogic page size – 15M
  - Classification of sub-LUN as high, medium, and low I/O
  - XVS moves data within tier based on workload class
  - Data movement done in the background
    - May take hours for movement depending on the workload
  - XVS reserves space in SSD as write cache
  - Recommended by Dell for VDI environments

# Dell EqualLogic AutoTiering

- EqualLogic volume tiering
  - Three tiering methods
    - Automatic Performance Load Balancing – default method
      - Monitors for Random I/O's
      - Tiers data within a pool for the entire volume
      - Can move the data to volume within pool that is configured for RAID 10
    - RAID Preference Tiering
      - Override of automatic method
      - Volume is associated with a RAID preference
      - Must have arrays of different RAID type
      - Could have volume moved to non-preferential RAID type
      - No automatic movement but automatic targeting by RAID level

# **Dell EqualLogic AutoTiering**

- Pool-based Tiering
  - Storage pool is defined as a tier
  - Volumes assigned to pool at creation
  - A tier (storage pool) could have different device types: SSD, 15K SAS, 10K SAS, SATA
  - A tier (storage pool) can have members (arrays) with different RAID protection
  - Manually controlled tiering

# **Dell EqualLogic Messaging**

- Intelligent Tiering
  - Part of storage consolidation solution
    - Separate storage pools in a SAN (with different devices)
    - Movement between storage pools is EqualLogic tiering
    - Tied to automatic load balancing across arrays

## **EMC VNX – FAST VP**

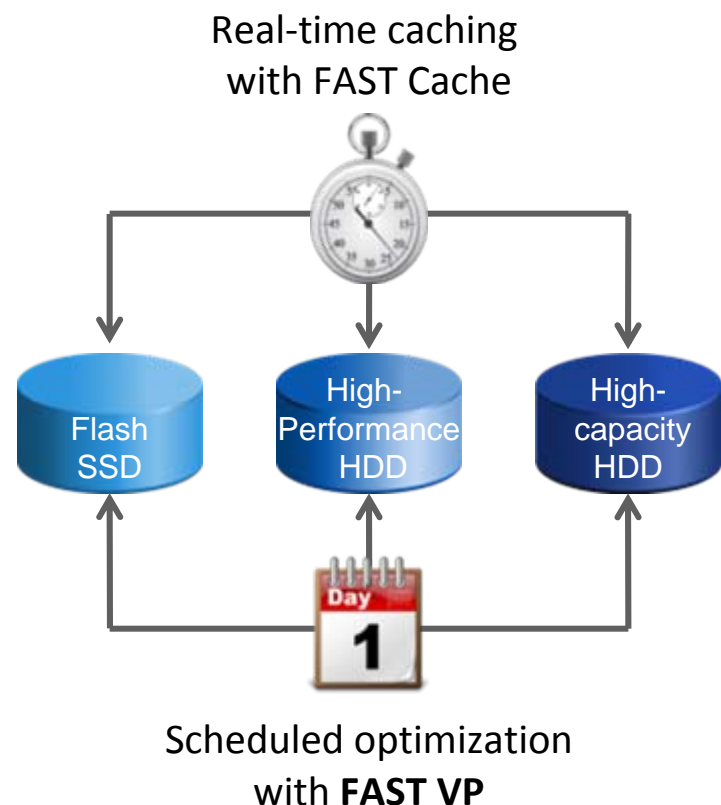
- Sub Volume – Sub-LUN
  - Supported with VNX
  - Supported on CX4 with Flare 30 (Shipped in August)
  - Supported on Celerra NS with DART v 6.0 (Shipped in August)
- Architecture of FAST VP
  - Requires Flare 30 (also required for Unisphere)
  - Below Flare 30: no auto load balancing of a LUN within a Thin Pool
  - Works on 1 GB slices – sub-LUN granularity
  - Supports up to 3 Tiers (defined by RAID level and / or Drive type)
  - Auto tiering works within a virtual pool

## **EMC VNX – FAST VP Operation**

- **Statistics collection—cumulative I/O history (reads and writes)**
  - Weights recent I/O history above longer-term I/O history
  - Maintains relative ranking of all data in pool, based on tier preference and I/O history:
    - Highest tier preference and high activity level get highest priority
    - Highest tier preference with less activity level get next highest priority
    - No tier preference (Auto-Tier) with high activity level gets next highest priority
- **On pool creation, detected in next poll cycle for inclusion in statistics collection**
  - LUNs created in pool are detected in next poll cycle for inclusion in statistics collection
  - Poll occurs every hour
  - Relocation estimate—“amount to move up/down”—updated every hour
- **Tier utilization**
  - Algorithm attempts to gain greatest utility from highest tiers
    - Data is demoted as space is needed in top tiers

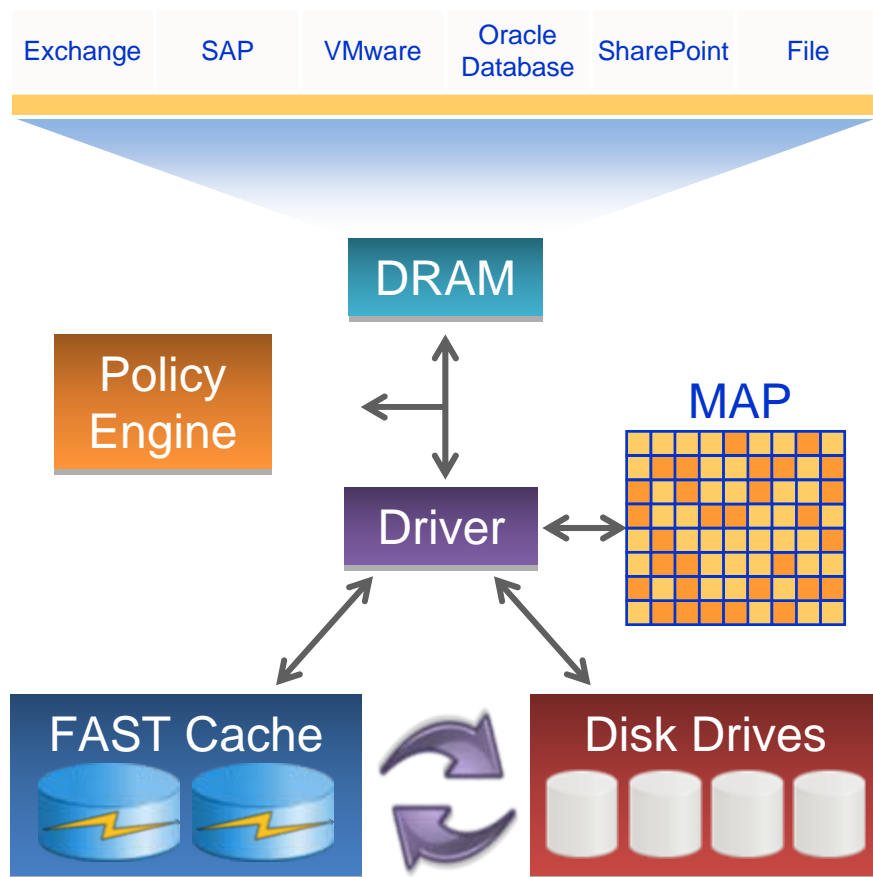
## EMC – FAST Cache

- FAST Cache
  - Supported on VNX, CLARiiON CX4 and Celerra NS w/ Unisphere
    - VNX support for both block and file
  - Supports up to 2TB SSD mirrored
  - Uses SSD as an extension of Cache
  - Supports read and write operations
  - Applies to classic LUNs and pool LUNs
    - Thick and thin provisioned LUNs
  - Unisphere setup with two clicks



Source: EMC

# EMC – FAST Cache Operation



- 1 Page requests satisfied from DRAM if available
- 2 If not, FAST Cache driver checks map to determine where page is located
- 3 Page request satisfied from disk drive if not in FAST Cache
- 4 Policy Engine promotes a page to FAST Cache if it is being used frequently
- 5 Subsequent requests for this page satisfied from FAST Cache
- 6 Dirty pages are copied back to disk drives as background activity

Source: EMC

# **EMC VNX FAST Messaging**

- Performance
  - Movement of active data for best performance
- Efficiency
  - Utilize high performance drives more effectively
- Lower cost
  - Fewer higher performance drives needed

# **EMC VMAX – FAST**

## **Fully Automated Storage Tiering**

- FAST Architecture
  - Full Volume (LUN) tiering
  - Supported with Enginuity 5874
  - Management is through GUI (SMC) or cli (SYMCLI) with FAST provisioning wizard
  - FAST v1 uses Symmetrix Optimizer
  - Analyze phase
  - Movement phase – swaps or moves between tiers are either automatic or user approved
  - Symmetrix Performance Analyzer (SPA) provides advanced monitoring for FAST environments

## **EMC VMAX – FAST VP Virtual Pools**

- Tiering at the Sub-LUN level
  - Virtually provisioned (thin provisioned) volumes
  - Storage groups of thin provisioned volumes
  - Extents size = 768K, extent group = 7.68MB
- Support 3 tiers
  - SSD (called EFD – Enterprise Flash Drives)
  - FC disks – 10K rpm or 15K rpm
  - SATA disk
- Requires Enginuity 5875

## **EMC VMAX – FAST VP**

- Two modes
  - Automatic – analysis and movement is ongoing
  - Off – performance data collected, no movement
- Initial release does not support CKD volumes
- Interoperates with other advanced features
  - SRDF, TimeFinder, Open Replicator, Virtual Provisioning, Space Reclamation, Symmetrix Optimizer, Data-At-Rest Encryption
- Both FAST and FAST VP can be simultaneously active

## **EMC VMAX – FAST VP**

- Two operational elements
  - Symmetrix Enginuity embedded software
  - FAST engine runs on the service processor
- Two tiering algorithms
  - Intelligent Tiering Algorithm
    - Collects I/O activity information and analyzes
    - Moves data based on performance metrics and tiering policies
  - Allocation Compliance Algorithm
    - Monitors allocated capacity in storage group
    - Moves data based on policies for capacity per group

# Intelligent Tiering Algorithm

**Service Processor  
FAST Controller**

**Symmetrix Enginuity**

**Performance  
Data Analysis**

Calculations to  
get move  
thresholds

- Runs continuously
- Can operate for 4 days without service processor

**Intelligent  
Tiering  
Algorithm**

Extents  
To move

**Data  
Movement  
Engine**

**Sub-LUN  
Performance  
Data Collection**

- Metrics Collected:**
- Read miss count
  - Write count
  - Sequential read prefetch

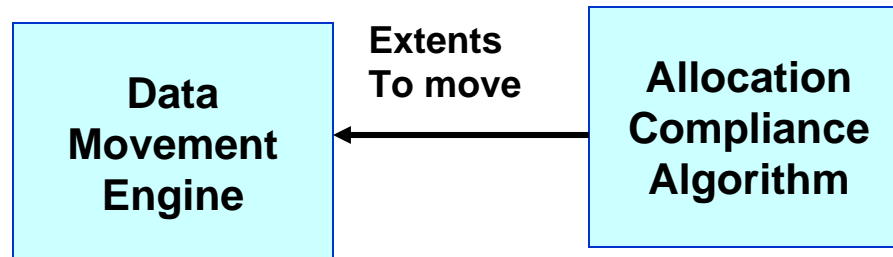
# Allocation Compliance Algorithm

## Symmetrix Engenuity

---

### Reasons for exceeding policy for Allocated capacity in storage group

- Adding thin LUNs with previous capacity
- Changes made to tiering policy
- Manual migrations
- New extent allocations for thinly provisioned LUNs as additional data is written



## **FAST VP Data Movement**

- Moves extents between pools
- New allocations for thin LUNs remain with the assigned pool for that LUN
- Movement is only active when
  - In automatic mode
  - During data movement window
  - Volume not pinned
  - Available space in target pool
- Max size request for single data movement is 10GB
  - Can be reduced if more than one move request
  - Size would be 10GB/# of requests

# FAST VP Management

- Controls
  - Wizard driven GUI set up for FAST VP set up and operation
  - FAST VP relocation rate can be set from 1 (fastest) to 10 (slow) to balance performance vs. speed at a system level
  - VLUN QOS can be used to control relocation rate at a volume level
  - Storage Group Priority can be used to manage migration rates at an application level
  - Pool reserved capacity: space reserved for accomplishing data movement. Default of 10% with range of 1% to 80% settable.
  - Up to 128 time windows can be defined for analysis and data movement periods.

# FAST VP Management

- Controls
  - Performance time window: when performance data is to be collected and not to be collected.

The screenshot shows the 'FAST - Settings (000192602812)' window with three tabs: 'General', 'Performance Time Window', and 'Move Time Window'. The 'Performance Time Window' tab is active, displaying a table with the following data:

Name	Include	Start Time (EDT)	End time (EDT)	Recurrence	Repeat
SAPPerfWindow	Yes	Nov-12-2010 19:00	Dec-31-2020 00:00	Once	NA
The Default Time Window	Yes	N/A	N/A	Once	NA

Below the table is a scrollable area with a scrollbar. To the right of the table are buttons for 'New', 'Edit', 'Delete', and two arrow buttons (up and down). Below the scrollable area is a 'Composite Time Windows' button.

Source: EMC

# FAST VP Management

- Controls
  - Data movement window: date and time ranges for when movement is allowed or not allowed.

FAST - Settings ( 000192602812 )

General Performance Time Window Move Time Window

Time Window Mode  FAST DP Move Time Windows  FAST VP Move Time Windows

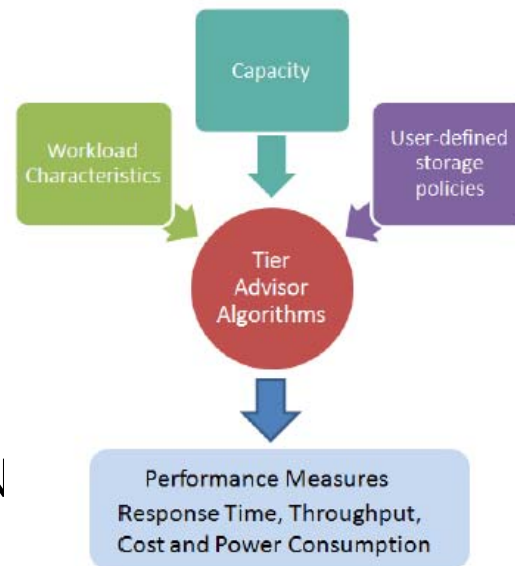
Name	Allow	Provisioning	Start Time (EDT)	End time (EDT)	Recurr
SAPVPMoveWindow	Yes	FAST VP	Nov-15-2010 16:30	Dec-31-2020 00:00	Once
The Default Time Window	No	FAST VP	N/A	N/A	Once

Composite Time Windows

Source: EMC

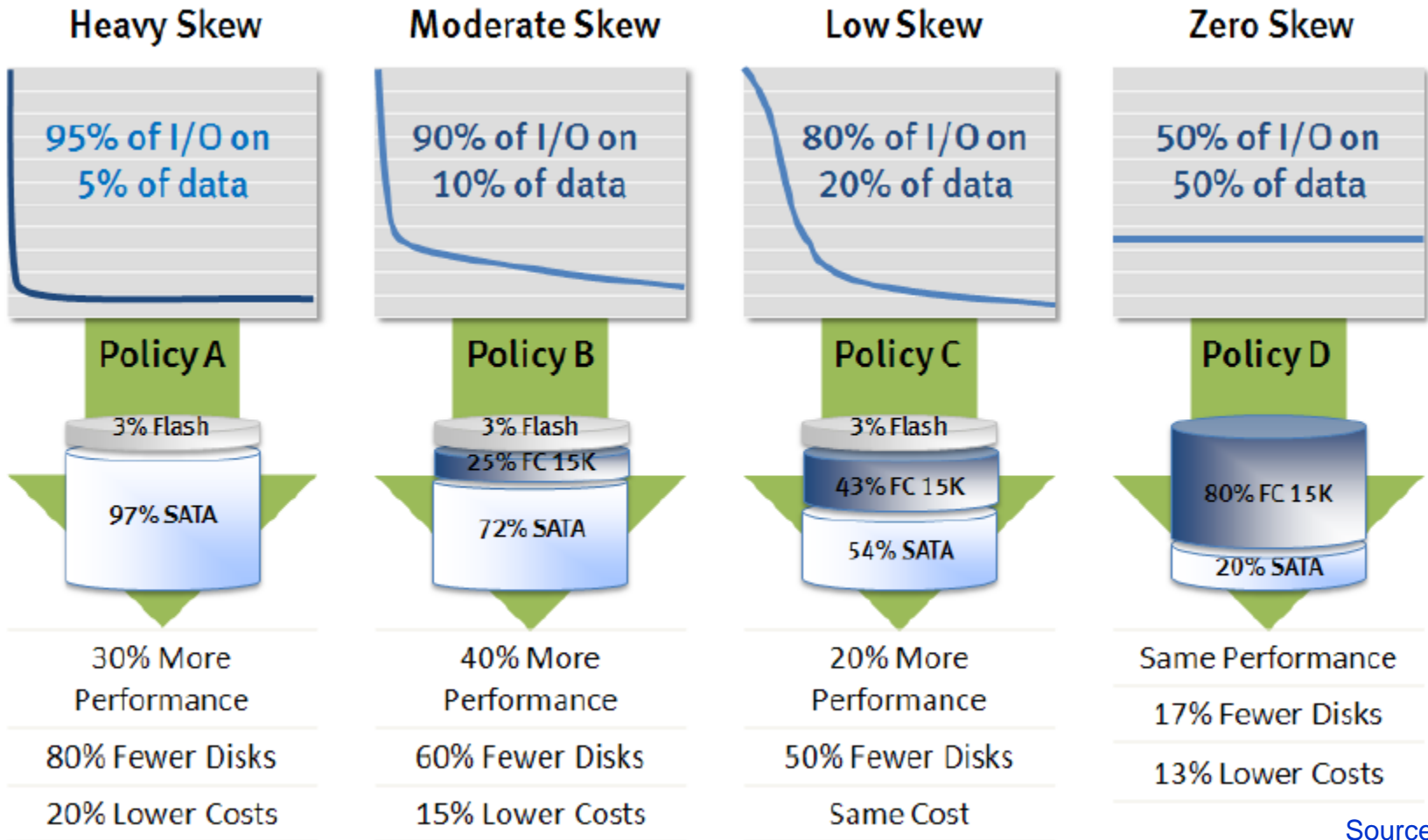
## FAST VP Tier Advisor

- Tier Advisor is a modeling tool which provides:
  - Ability to create an optimal mix of drive types (SSD, FC, SATA) for both cost and performance
  - Helps plan new FAST implementations
  - Supports both VMAX and VNX/CLARiiON arrays
  - Provides FAST VP pool estimates
  - Supports importing configuration and performance data from one or multiple storage arrays for analysis and consolidation



Source: EMC

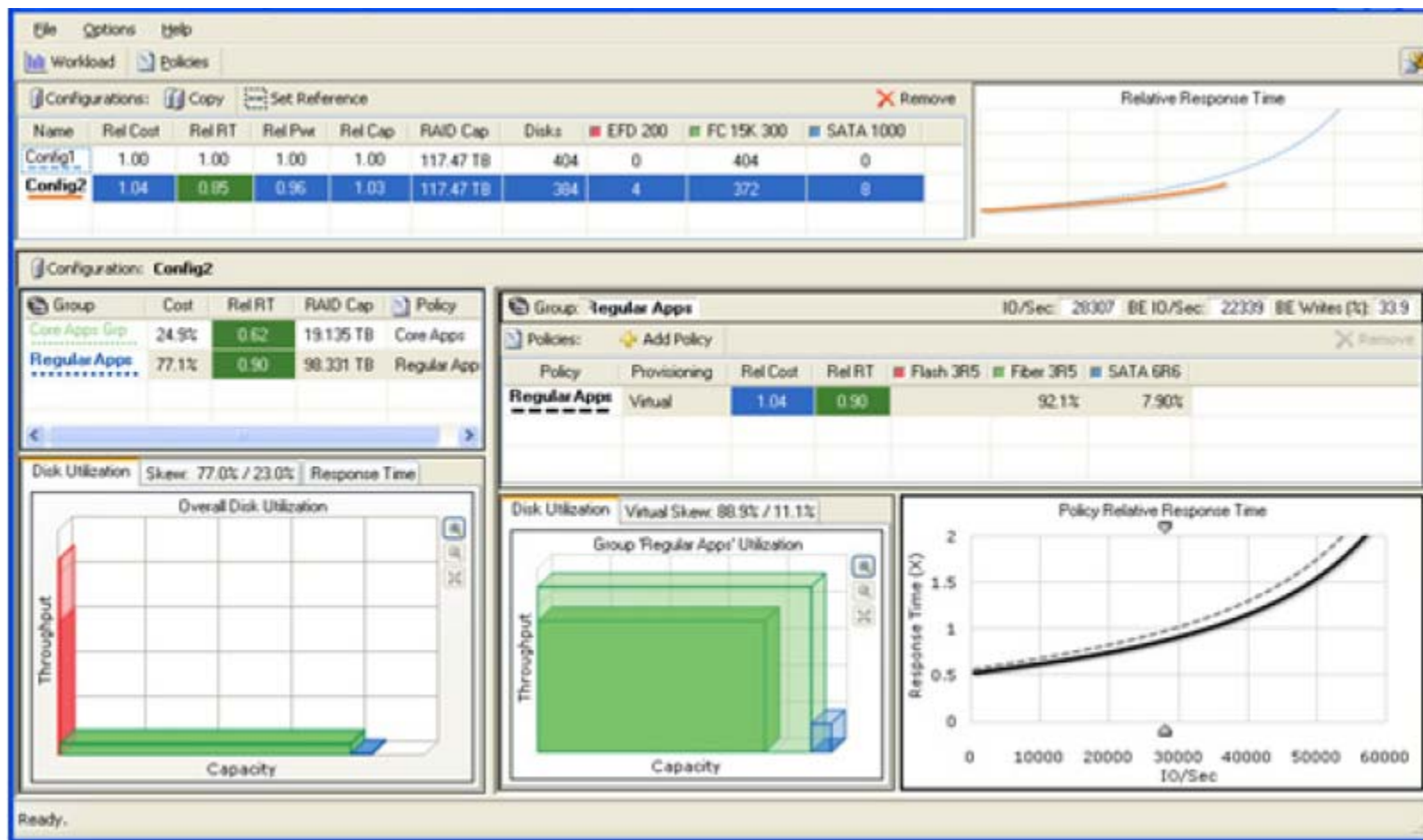
# FAST VP Tier Advisor



Source: EMC

**Workload skews compared to a 100% Fibre Channel configuration**

# FAST VP Tier Advisor



Sample output from Tier Advisor

Source: EMC

# **EMC VMAX FAST Messaging**

- Performance Optimized
  - Support for multiple drive options
  - Real-time: Adapts to business needs
- Reduces system costs
  - Fewer higher performance drives needed
  - Lower power and cooling with SSD usage

# HP 3PAR – Adaptive Optimization

- Architecture
  - Based on Autonomic Storage Tiering
  - Supports LUN Tiering (Dynamic Optimization)
  - Also Sub-LUN (Adaptive Optimization)
  - Uses a 128 MB “mini” chunk (1/2 of 256 MB chunklets)
  - Up to 3 Tiers, defined by
    - Drive type, RAID level, stripe width and radial placement of data
  - May be enabled or disabled per volume
  - Licensed feature
  - Also requires 3PAR Reporter
- Products
  - All F and T
  - Utilizes 100 GB SSD's

# HP 3PAR Dynamic Tiering Messaging

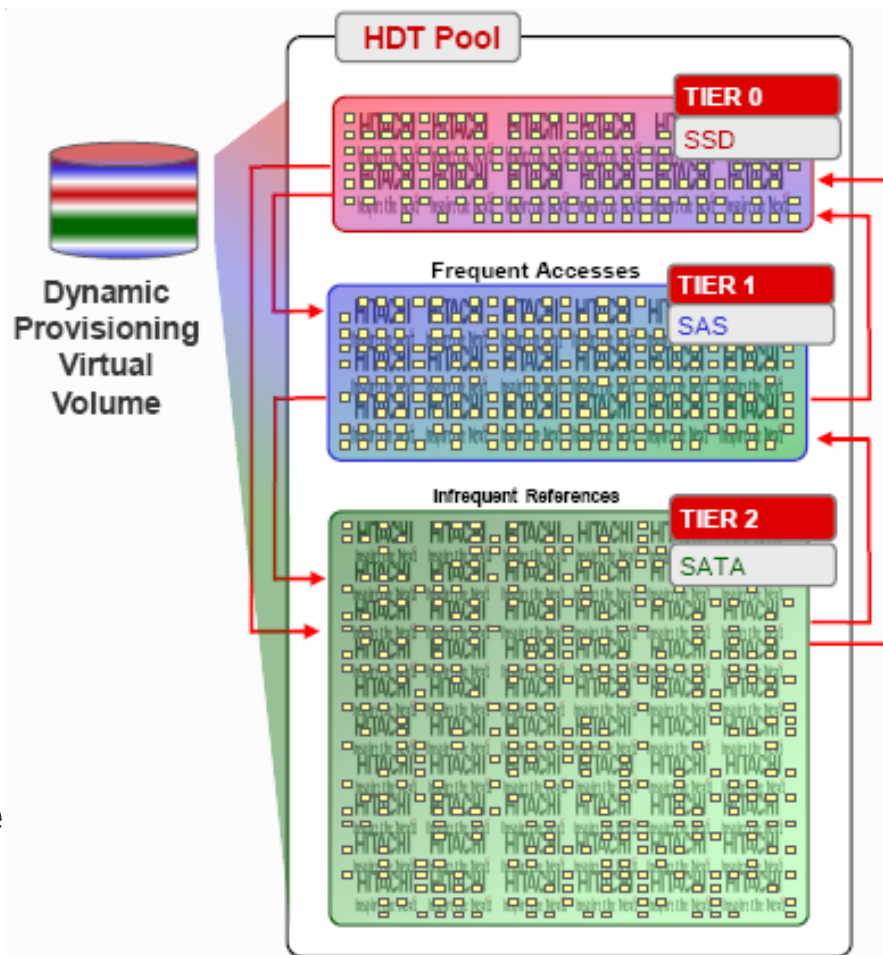
- References two types:
  - Adaptive optimization (sub-LUN tiering)
    - QoS management
    - Cost reduction – minimize number of SSDs required to average 30% lower costs
    - Optimized performance
    - Risk avoidance for IT managers – automatically handle changing workloads
  - Dynamic optimization (volume level tiering)
    - QoS management on volume basis
    - Policy advisor software – to report and move data

## **Hitachi VSP – Dynamic Tiering**

- VSP supports “Dynamic Tiering” - Sub-LUN
  - SSD, SAS, and SATA disks per pool
  - Only internal (captive) storage currently
- Architecture
  - Uses page tiering with page chunks of 42 MB
  - New data staged to highest tier
  - Supports 3 tiers
    - Based on 3 criteria: Media Type, Drive Speed and RAID level
  - Volumes may be included, or excluded
  - Movement can be controlled by time of day
  - Movement cycle is between 1 – 24 hours

# Hitachi Page Level Tiering

- Tiering Operation
  - Different tiers of storage are now in one pool of pages
  - Data is written to the highest performance tier first
  - As it becomes less active it migrates to lower level tiers
  - If activity increases it will be promoted back to a higher tier
  - Since 20% of data accounts for 80% of the activity, only the active part of a volume will reside on the higher performance tiers



Source: HDS

# **HDS Dynamic Tiering Messaging**

- Improved performance
  - Highest efficiency usage of storage resources
- Reduced costs
  - Fewer high performance drives required
  - Maximize SSD effectiveness
  - Eliminates data classification requirements
  - Reduces power, space, cooling requirements

## **IBM Easy Tier**

- Architecture – DS8000 and V7000
  - Extension of caching algorithms
  - Manual or Automatic modes (only Automatic is Sub-volume)
  - Different storage cannot be mixed in extent pool, except w/ Easy Tier
  - Granularity = extent size
  - No controls per volume, either “on” or “off”
  - Two tiers only
- Analysis of data access, migration plan is created & executed
- Every 24 hours, a new migration plan is created

## **IBM Easy Tier - continued**

- Nomenclature
  - Array site
    - Group of 8 drive modules for DS8000, up to 24 for V7000
    - Each drive in array must be same type (and speed)
  - Array
    - Created from one array site during configuration
    - FC or SAS disks – RAID 5, 6, 10
    - SSD's – RAID 5
    - SATA – RAID 6, 10
  - Ranks
    - Created from one array site
    - Name of rank assigned by DS Storage Manager
    - DS8000 - A rank (therefore Array) can be formatted into either Fixed Block or Count-Key-Data
      - FB extent size = 1GB
      - CKD extent size = 1113 cylinders (.94GB)
    - V7000 - Size is variable, from 16 MB to 8 GB, default size of 256 MB

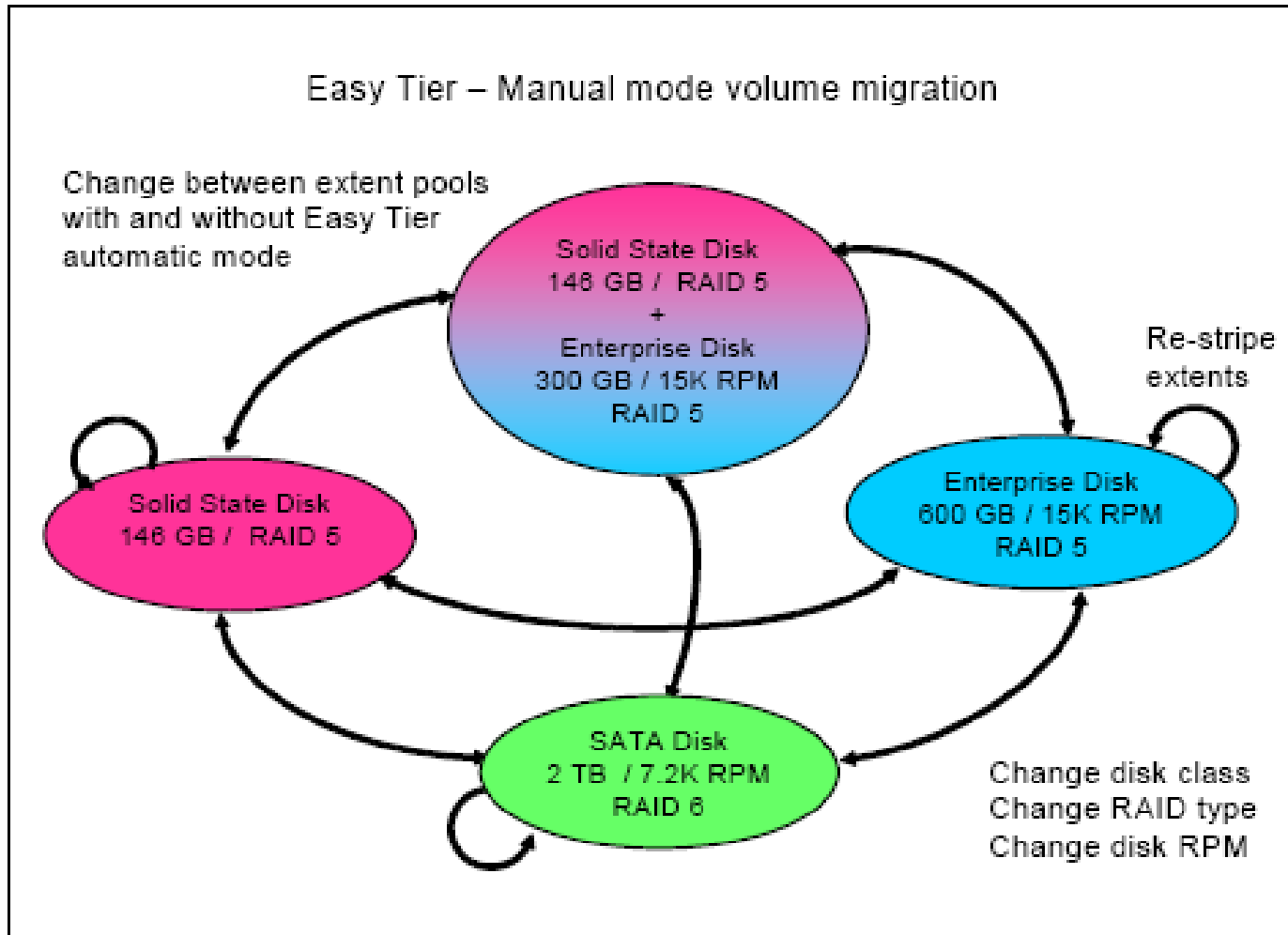
# IBM Easy Tier Messaging

- Efficient
  - Effective use of SSD capacity
  - Obtain the most performance from a storage system
- Economical
  - Reduce the number of SSDs
- Intelligent
  - Learning mode to intelligently move data – automatically and non-intrusively

# IBM Easy Tier

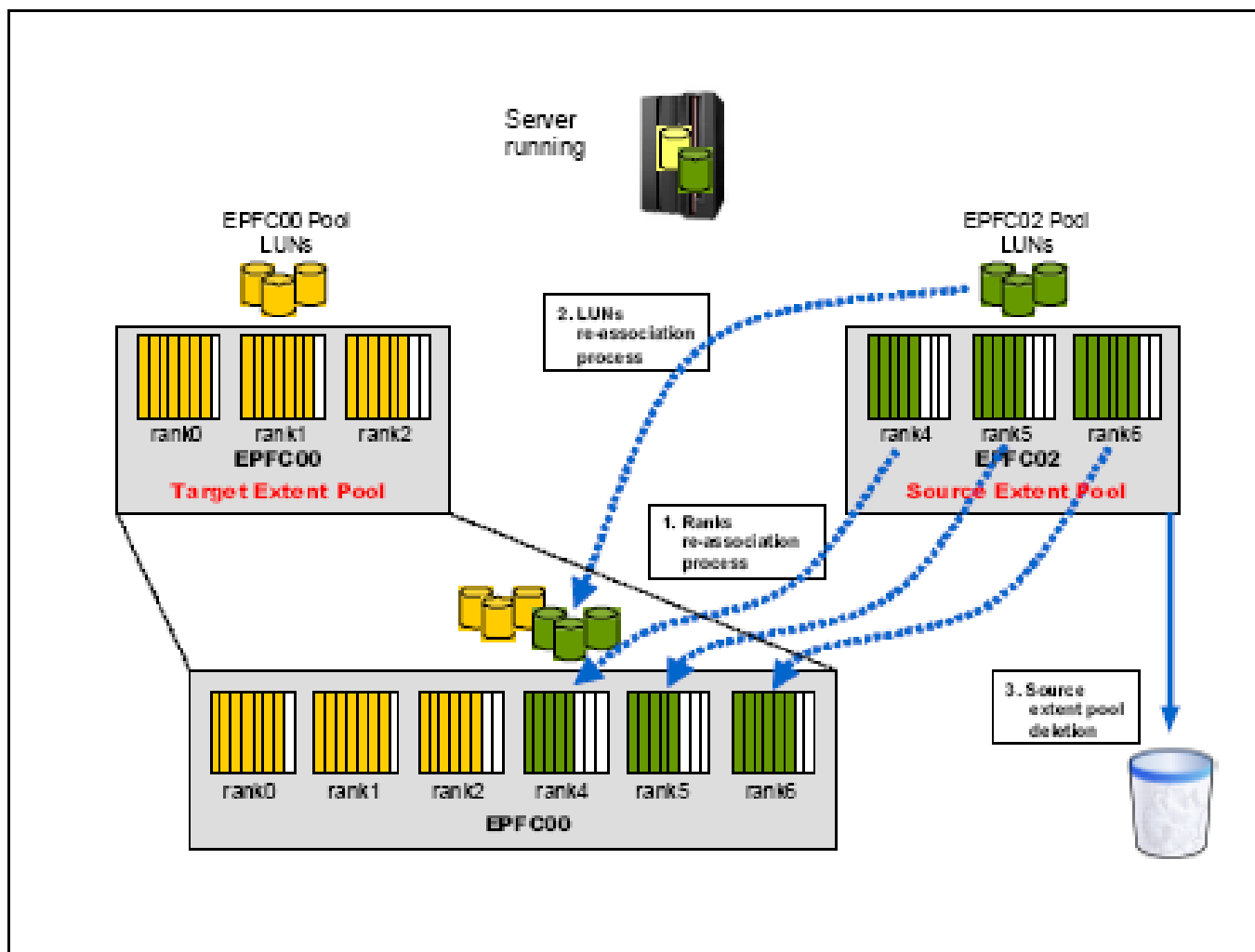
- Two Operational Modes
  - Automatic mode
    - Manages data in extent pools that contain two storage technologies
    - Can be sub-LUN/Volume
    - Can selective choose volumes for automatic mode, all, or none
  - Manual mode
    - Manually initiate action to relocate data at volume level
    - Dynamic volume relocation – data moved from one extent pool to another
    - Dynamic extent pool merge – merges one extent pool with another

# Easy Tier Manual Mode Migration



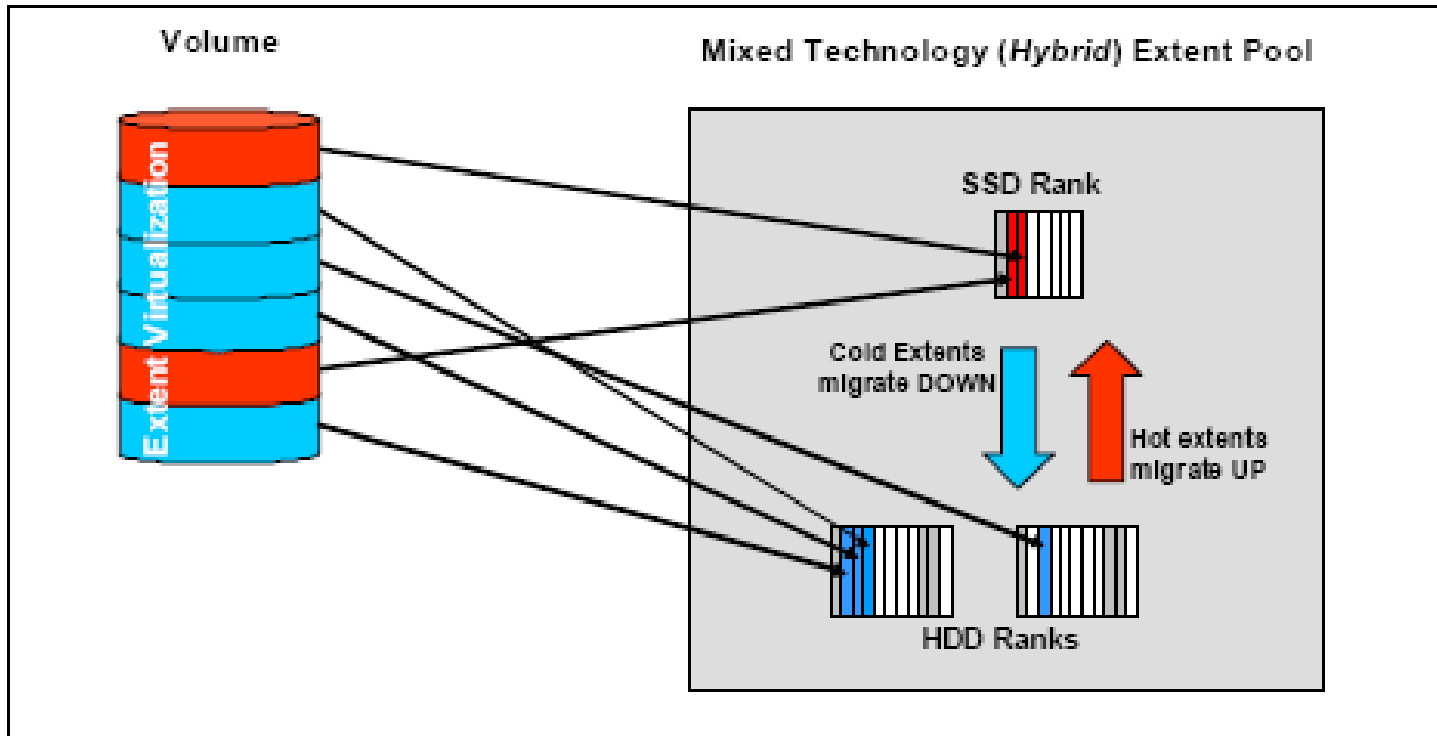
Source: IBM

# Easy Tier Manual Mode Merge



Source: IBM

# Easy Tier Automatic Mode



# Easy Tier Storage Advisor

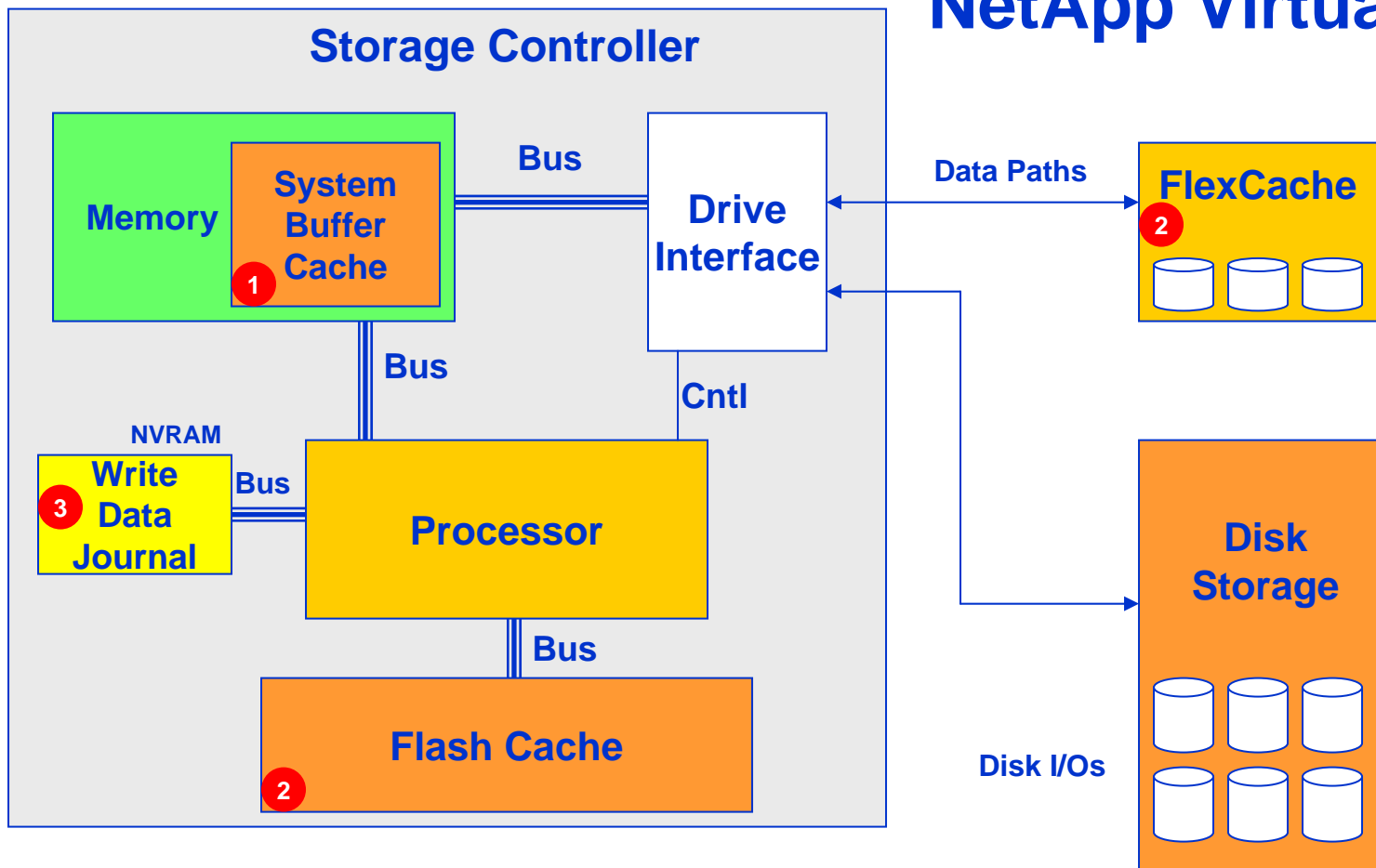
- Tool for monitoring and managing Easy Tier
  - Graphical representation of performance data
  - Collected over a 24 hour cycle
  - Application that runs on a Windows system
    - Summary report extracted from storage system and hardware management console
    - Report is a heat index for data

# NetApp Virtual Tiering

- Avoids migration of data between different drive elements
  - Does not consume disk I/O's
  - Does not consume internal controller bandwidth / data paths
- Uses multi-level read cache architecture
  - Buffer Cache – system memory
  - Flash Cache – optional read cache using Flash module (formerly called PAM-II)
  - FlexCache – optional use of disks for read caching

See Evaluator Group analysis of NetApp Virtual Tiering for more information

# NetApp Virtual Tiering



- 1 - Read and Write data caching
- 2 - Read data caching
- 3 - Write data journal for cached data

# NetApp Virtual Tier Messaging

- Performance
  - No resources consumed for moving data between tiers
  - Intelligent placement of data
- Caching
  - Use internal bus for memory I/O

# Xiotech Hybrid ISE Tiering

- Architecture
  - Hybrid ISE supports both SSD and disk in sealed modules
  - Continuous Adaptive Data Placement to control placement of data across devices – can be SSD or disk based
    - Includes intelligent tiering
    - Fully automated – no monitoring or changes required
    - Target is to place highly active random I/O in SSD
  - Sub-LUN data movement based on activity
    - Continuous monitoring of accesses using weighted I/O counts
    - Starts within one minute of I/O
    - Granularity of movement is 240MB (called a sheet)
    - Takes less than 4% bandwidth of ISE module
    - Maximizing I/O's may take several hours
  - Two tiers within Hybrid module

# **Xiotech Tiering Messaging**

- True automation
  - Automation with initial data placement controls and movement based on type of activity
  - No intervention or setting required
- Non-impacting
  - Done at ISE module level, not storage system controller
  - Original placement decisions minimized movement needed
  - Small overhead required
- Intelligent
  - Uses ROI calculation on weighted I/O counts to make decisions
  - History of migrations kept to maximize migration value

## Questions?

- Evaluator Group Website
  - <http://www.evaluatorgroup.com>
- Evaluator Group Contact Information
  - Phone: (303) 221-7867
  - Email: [info@evaluatorgroup.com](mailto:info@evaluatorgroup.com)



**Evaluator Group**