

A GPS Guide for Your Journey to Automated Storage Tiering

Tony Negro
EMC Corporation

Tuesday August 13, 2013
Session Number 14125



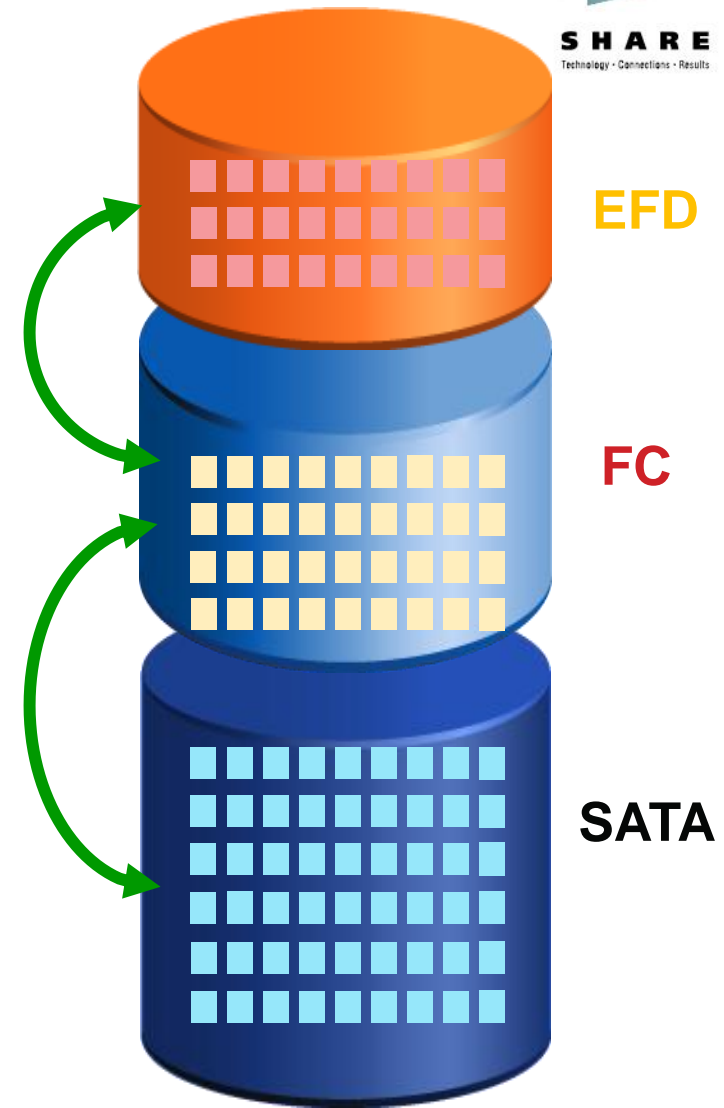
AGENDA

- Introduction (**Why make the journey**)
 - Virtual Provisioning for CKD devices
 - Fully Automated Storage Tiering for Virtual Pools (FAST VP)
- Implementation Planning (**preparing for the trip**)
 - A Starting point
 - Understanding workload Skew
- Configuration Best Practices (**avoiding traffic jams**)
- Migration Considerations (**navigation choices**)
 - Host based vs. Array based
 - Allocation by policy
 - Monitoring FAST VP
- Summary

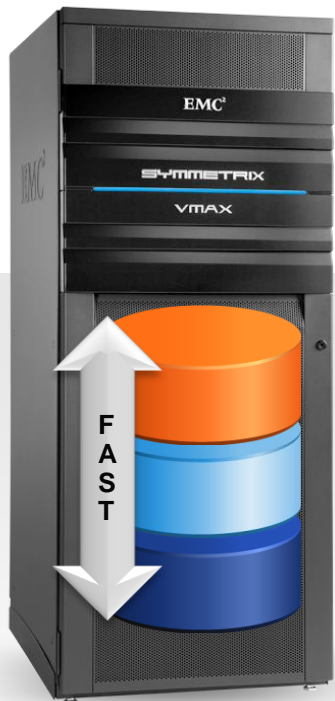


Introduction

- **FAST VP** – Fully Automated Storage Tiering for Virtual Pools
- Policy-based system that automatically optimizes application performance
- Promotes and demotes sub-LUN level data across storage tiers to achieve performance service levels and/or cost targets



Why FAST Adoption



Financial Services

56% LOWER
COST per ARRAY

Investment Banking

25% FASTER
RESPONSE TIME

Health Care
Provider **1.5X**
MORE
UTILIZATION

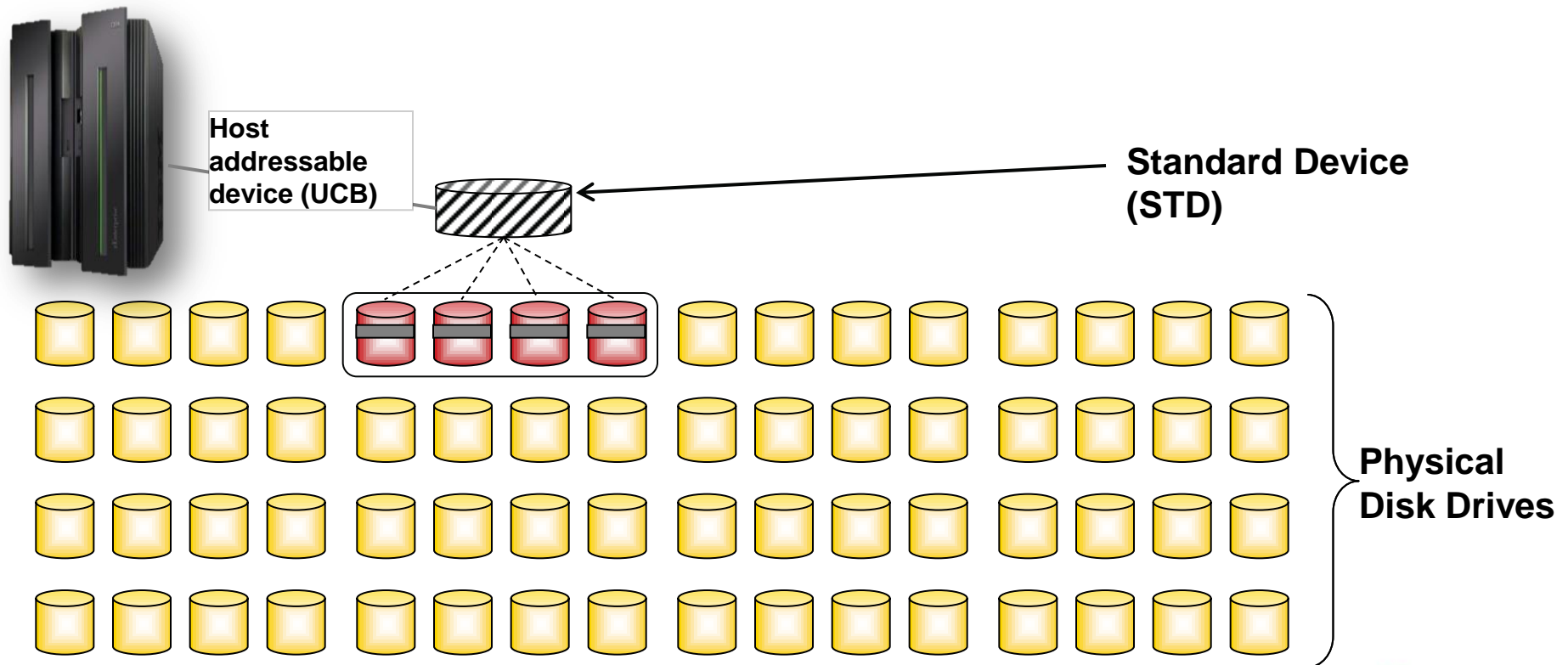
Insurance
40%
LOWER
ENVIRONMENTAL COST

Virtual Provisioning for CKD Devices

Overview

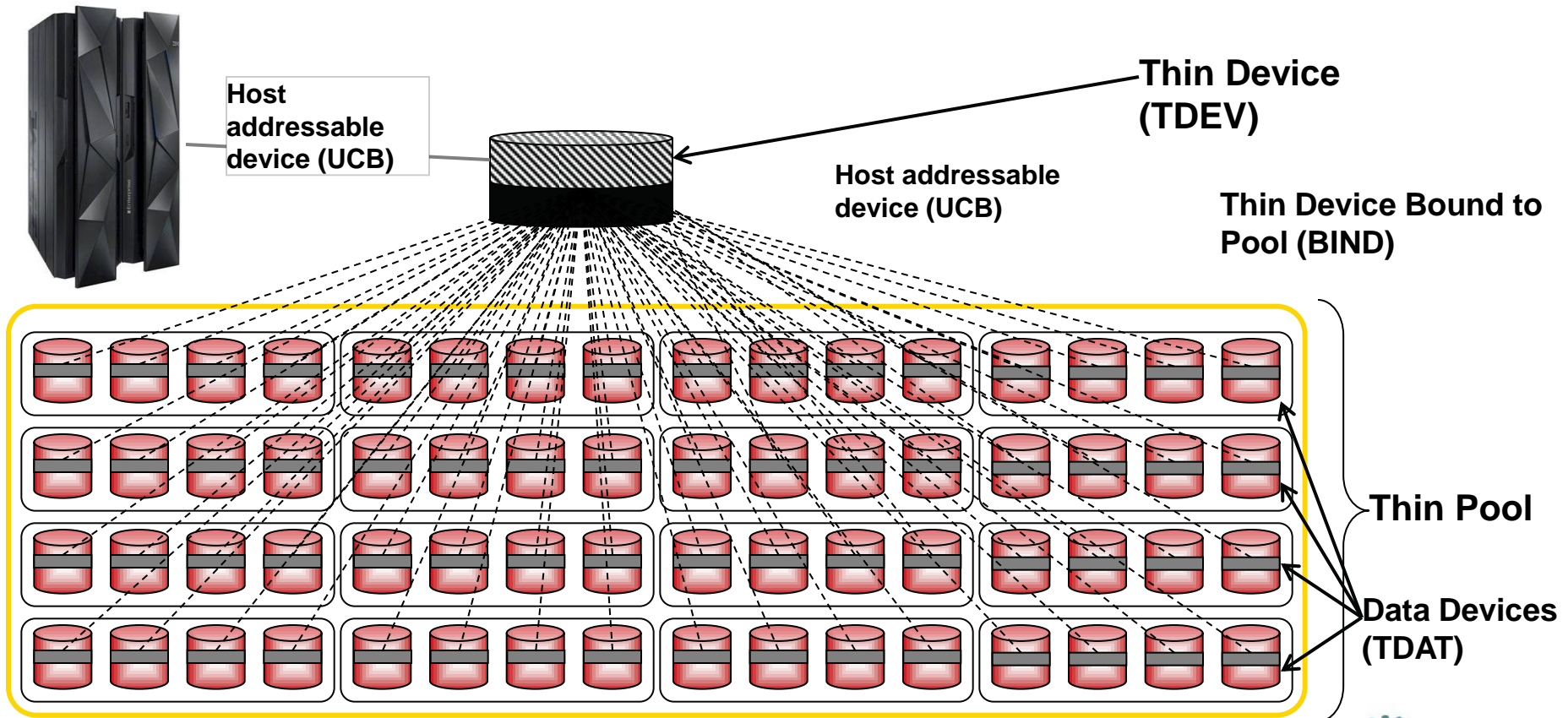
Standard Provisioning Concept

- Capacity for a CKD device is allocated from a group of physical disks
 - Example: RAID 5 with striped data + parity
- Workload is spread across physical disk of the RAID group



Virtual (thin) Provisioning Concept

- Storage capacity is structured in virtual pools
- Workload is spread across physical disk in the pool



Theory of Operation

- Two new types of devices are introduced:
 - **Thin device** – the host visible device which is configured to customer requirements.
 - Can be used with SRDF, TimeFinder and IBM Compatibility features
 - **Data device** – an internal device that contains storage space (Track Groups) that are associated with a thin Device
- Actual space from a data device is assigned to a thin device during the BIND operation or on demand when data is written

Over-subscription

- An over-subscribed thin pool has more total reported thin device capacity than is physically available
- Over-subscription can reduce consumed storage
 - Avoids over-allocation of physical storage to applications
 - Reduces the cost of storage, energy consumption and footprint
- Care must be taken to ensure that a pool full condition is not reached
- Pool consumption can be monitored with Solutions Enabler SYMCLI, Solutions Enabler Event Daemon, UniSphere for VMAX or with SCF Monitor

Pre-allocated Pool Space

CU#	SSID	TYPE	CYL	GB/CU	3390B
00	1000	3390-27	32760	6236.16	224
01	1001	3390-27	32760	6236.16	224
↓	↓	↓	↓	↓	↓
0B	100B	3390-27	32760	6236.16	224
0C	100C	3390-9	10017	2042.4	240
0D	100D	3390-3	3339	681.6	240
				71321.76	2944

Thin Devices (TDEV) = ~72TB



120%Over-subscribed Pool

CU#	SSID	TYPE	CYL	GB/CU	3390B
00	1000	3390-27	32760	6236.16	224
01	1001	3390-27	32760	6236.16	224
↓	↓	↓	↓	↓	↓
0B	100B	3390-27	32760	6236.16	224
0C	100C	3390-9	10017	2042.4	240
0D	100D	3390-3	3339	681.6	240
				71321.76	2944

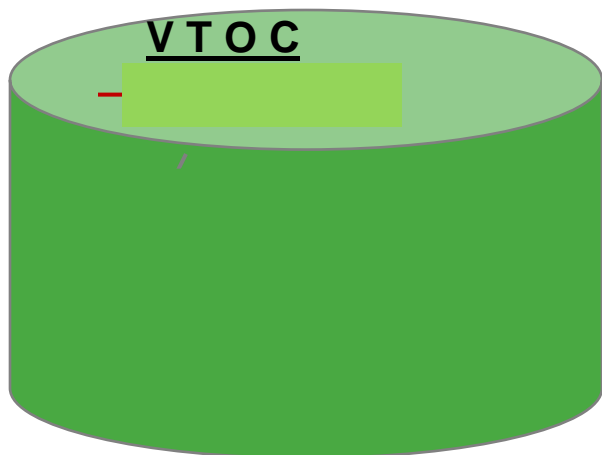
Thin Devices (TDEV) = ~72TB



Thin Reclaim Utility (TRU)

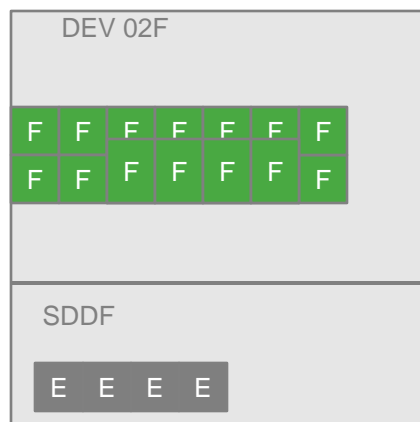
[1] TRU discovers and monitors thin devices that are not bound with PREALLOCATION /PERSIST attribute

DEV 02F



SCRATCH DSN=x.y.x

[2] EMC z/OS SCRATCH exit (TRU SCAN utility) identifies the tracks that are eligible to be reclaimed



F=Free A=Allocated E=Eligible

[3] TRU RECLAIM Utility periodically marks eligible tracks as “empty” by updating STD rec. 0 (*also can be done by batch job*)

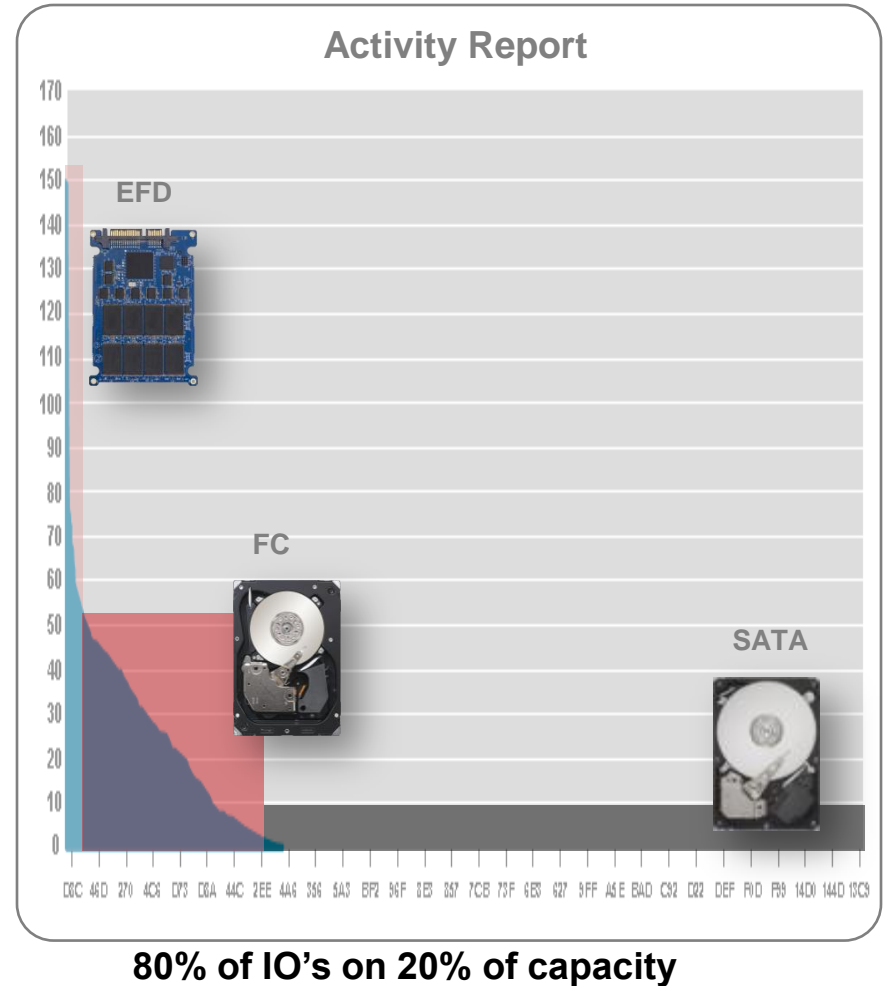
[4] Enginuity Zero Space Reclaim background process reclaims tracks with no user records (only STD 0)

Fully Automated Storage Tiering for Virtual Pools (FAST VP)

Overview

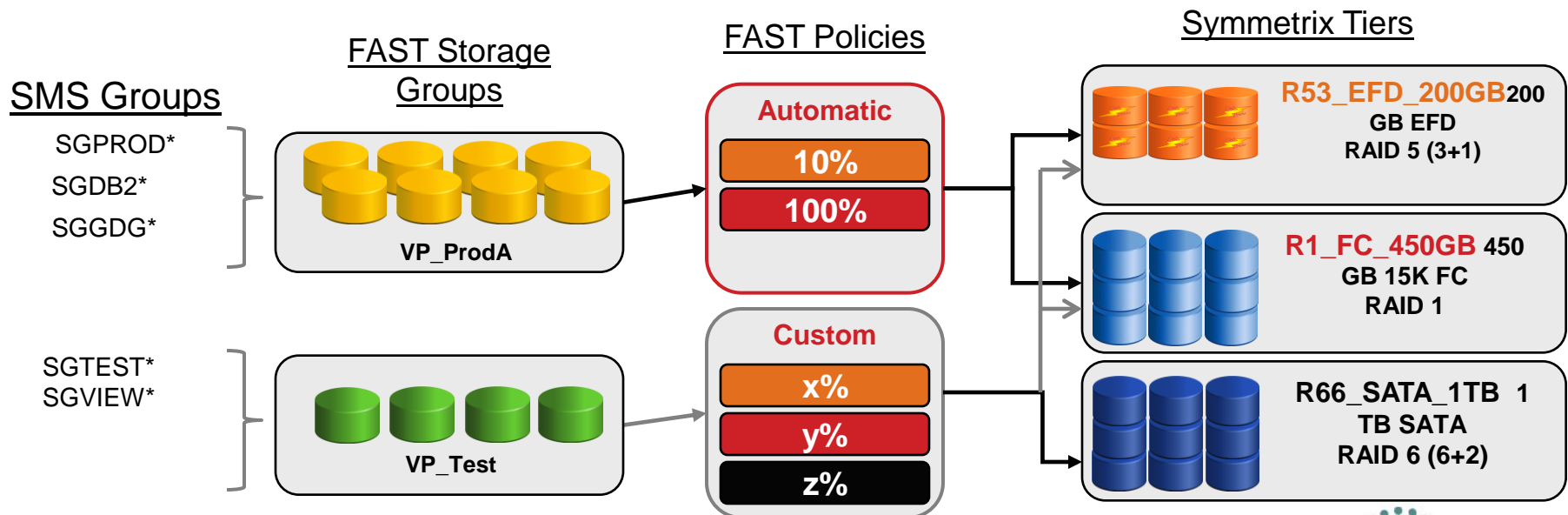
Basis for FAST

- With information growth trends, an ALL Fibre Channel configuration will:
 - Cost too much
 - Consume too much energy and space
- FAST VP helps by leveraging disk drive technologies
- What makes FAST work in real-world environments?
 - **Skew:** At any given time, only a small address range is active – the smaller the range, the better



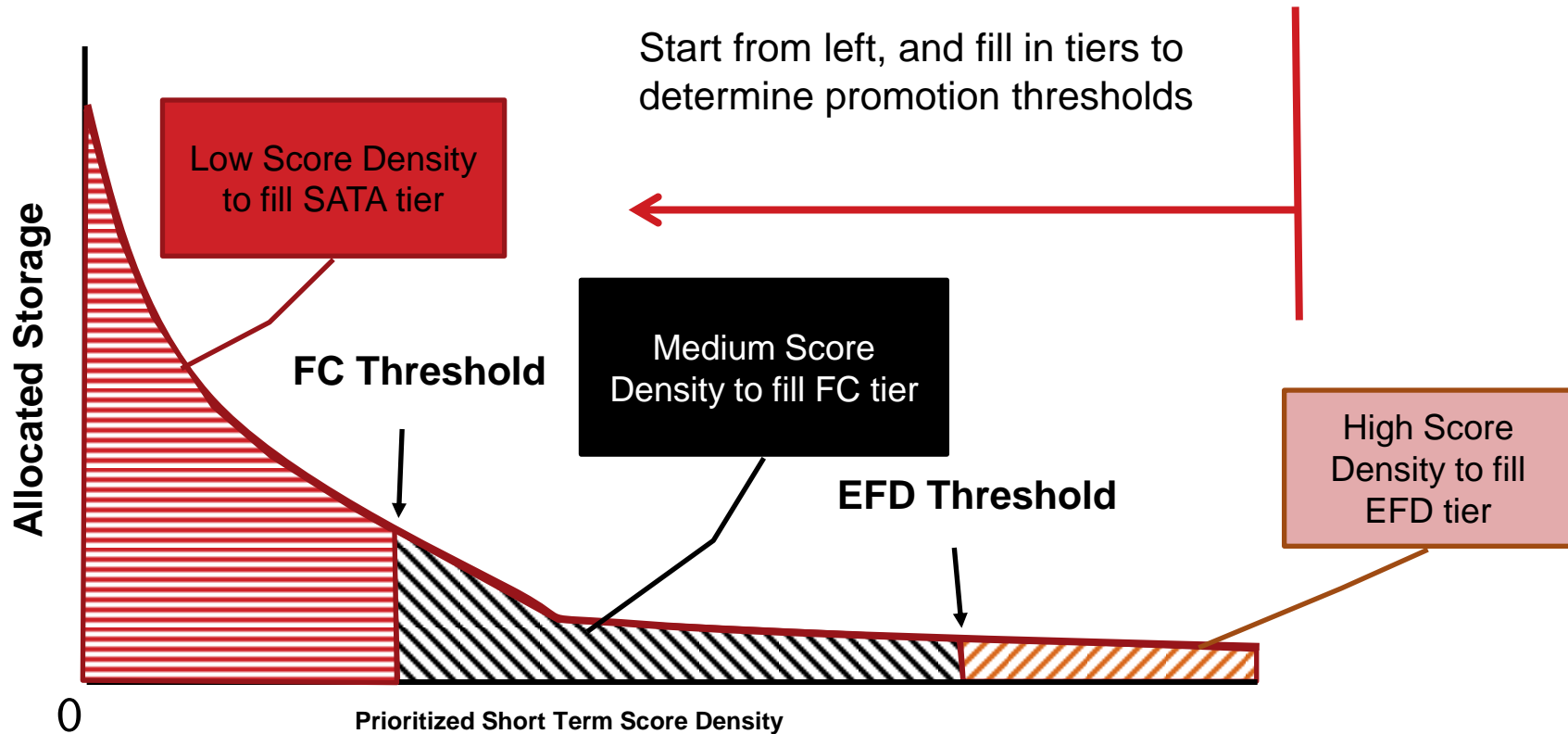
FAST Storage Elements

- Symmetrix Tier – a shared storage resource with common technologies (Disk Groups or Thin Pools)
- FAST Policy – manage Symmetrix Tiers to achieve service levels for one or more Storage Groups
- FAST Storage Group – logical grouping of devices for common management



FAST VP – Score Analysis

- Extents Group Sets are grouped in a histogram according to the Score





FAST VP Hierarchy

- **Extent Group Set**
 - 48 extent groups
 - 360 MB FBA / 320 MB CKD
 - **Statistics collection unit**
- **Extent Group**
 - 10 thin device extents
 - 7.5 MB FBA / 6.8 MB CKD
 - **Largest data movement unit**
- **Thin Device Extent (Track Group)**
 - VP allocation unit
 - 768 KB FBA / **680 KB CKD**
 - **Smallest data movement unit**

- I/O rates are continuously collected for each Extent Group Set
 - Read Miss (RM)
 - Write (W)
 - Prefetch (P)
- Rates are updated every 10 minutes changing the 'score' of the Extent Group Set

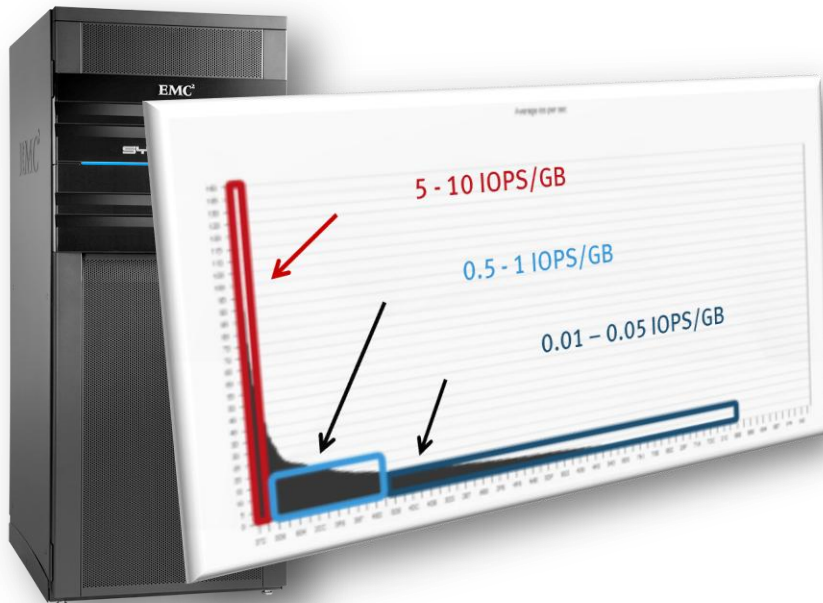
Implementation Planning

Preparing for the journey



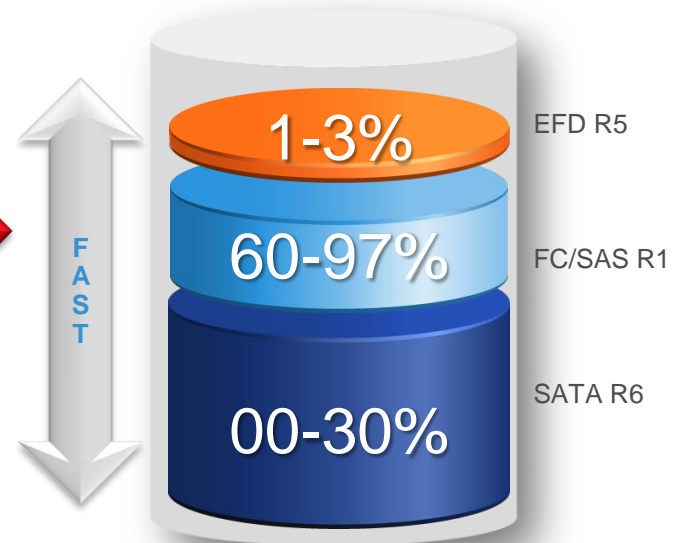
Configuration Starting Points

Typical Observed Workload Skew



85% or more of system IOPS from 15% or less of system capacity

Typical FAST Capacity Mix

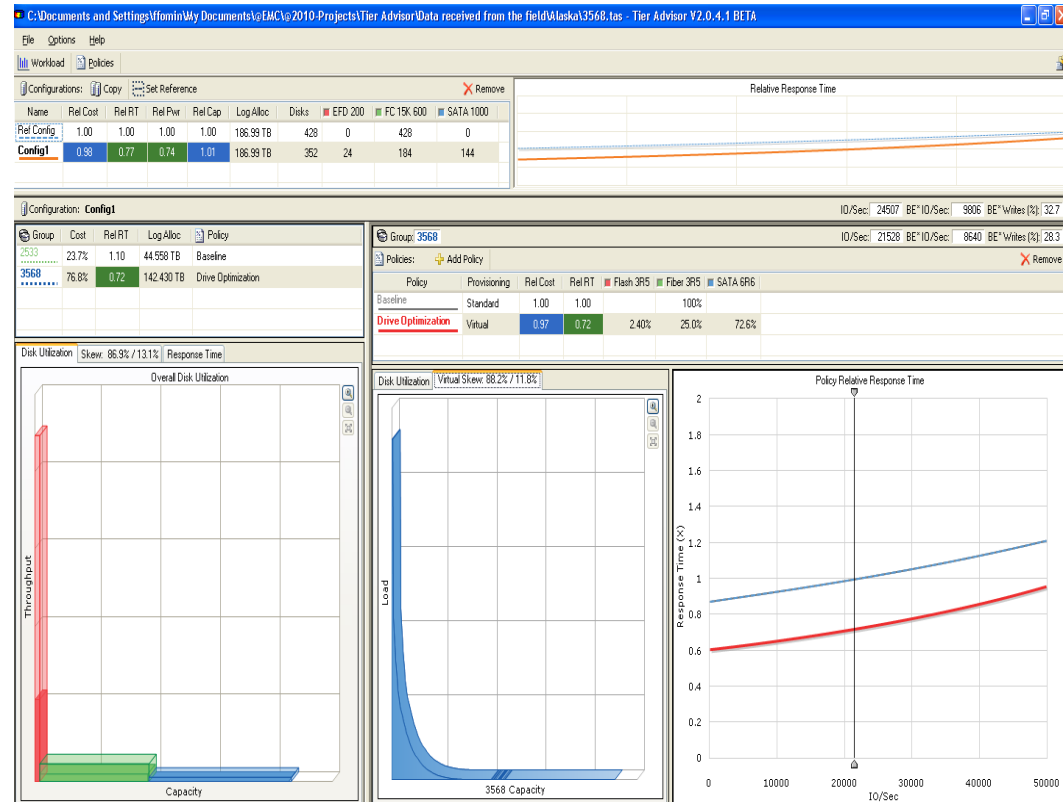


Use Tier Advisor to provide modeling based on actual workload skew

Tier Advisor

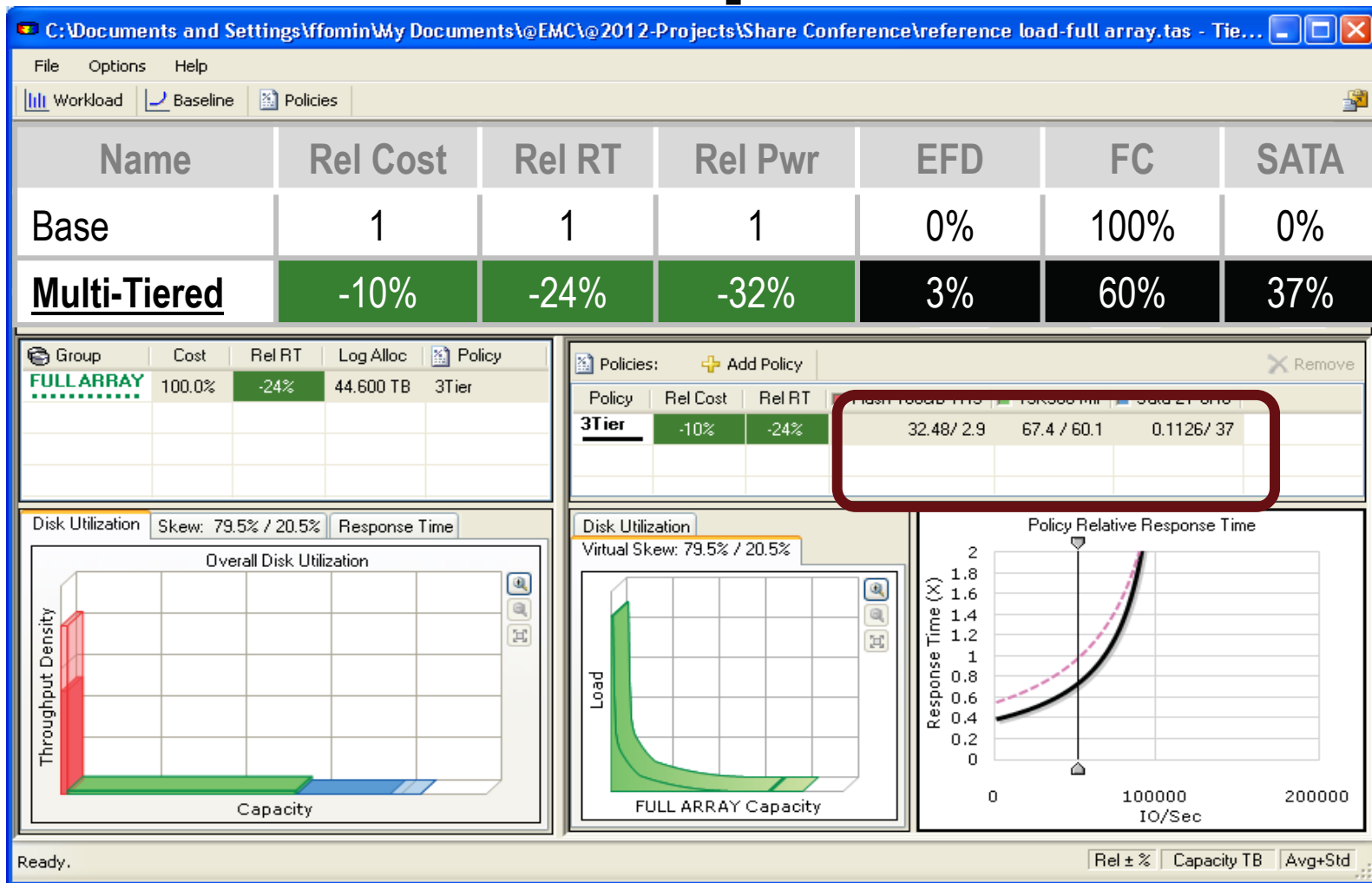
Disk modeling tool to:

- Identify Skew & Workload
- Find best price vs. performance combination
- Ensure that Tiers have the capabilities to support current load and growth



Tier Advisor is recommended for planning.
It is not required for FAST VP operations.

Potential for improvements



Workloads have different needs

C:\Documents and Settings\ffomin\My Documents\@EMC\@2012-Projects\Share Conference\reference load.tas - Tier Advisor ...

File Options Help

Workload Baseline Policies

Name	Rel Cost	Rel ST	Rel Pwr	EFD	FC	SATA
Base	1	1	1	0%	100%	0%
Multi-Tiered	-10%	-24%	-32%	3%	60%	37%

Group	Rel. RT	Capacity	Policy
DB High Prty	-26%	8 TB	3 Tiers
DB Very High	-48%	3 TB	EFD +FC
BATCH	-6%	17 TB	FC Only
Other	-10%	15TB	FC and SATA

FAST VP Time Windows

- Two time window types – **Performance** and **Data Movement**
 - Control what workload to include in analysis and when to allow FAST VP to move data
- Configure the **performance time windows** to target the specific workloads to optimize
 - Default is 24x7, but can be configured to exclude certain time periods
- Configure **data movement windows** to allow data movements all the time
 - At a minimum, configure for the same periods of time covered by the performance time windows

FAST VP Workload Analysis Period

- The **Workload Analysis Period** determines the degree to which FAST VP metrics are influenced by past host activity
- The higher the value, the more weight is assigned to less recent host activity
 - Leaves previously active data in a higher tier for as long as possible
- A lower value can allow FAST VP to be more aggressive in identifying data to be demoted
 - This may lead to greater amounts of data being moved between tiers
- Match the value to the performance time window to cover one calendar week

VP Allocation by FAST VP Policy

- When enabled, new allocations can be made in any pool contained within the associated FAST VP policy
 - Alleviates possibility of new allocation failing if chosen pool is full
- Enable **allocation by policy**
- FAST VP picks the best tier and pool
 - Based on FAST VP metrics and policies
 - Uses policy percentages and pool free space if no metrics

FAST VP Pool Reserved Capacity



- Reserves a percentage of a thin pool for new allocations
 - Reserved space excluded from FAST VP data movements
 - Set at system level for all pools, but can be overridden for individual pools
- Increasing **PRC** on a pool may cause FAST VP to move data out of that pool
 - Potential for creating additional space in a pool prior to a migration
- For pools with bound thin devices set the **PRC** based on the lowest allocation warning level for that thin pool
 - For pools with no bound thin devices, set the **PRC** to 1%

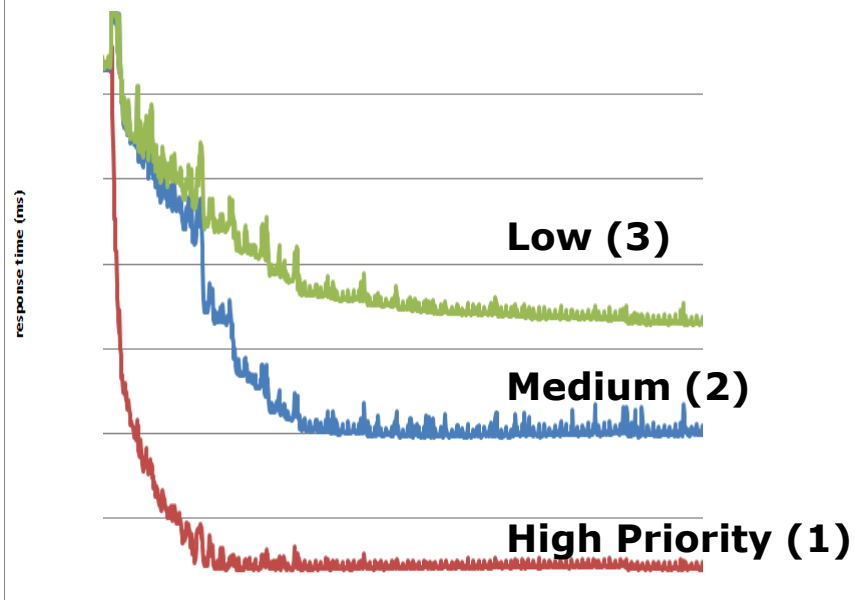
FAST VP SRDF Coordination

- Allows FAST VP promotion/demotion decisions on remote R2 devices to include local R1 performance metrics
 - R1 device read activity accounted for on R2 device
 - Potential to maintain more balanced configuration between R1 and R2 devices
 - R2 device data more appropriately allocated in case of failover
- Enable **SRDF Coordination** for SGs containing SRDF devices
 - Recommended for both SRDF/S and SRDF/A

Now Available! SRDF Coordination for Multi-site SRDF

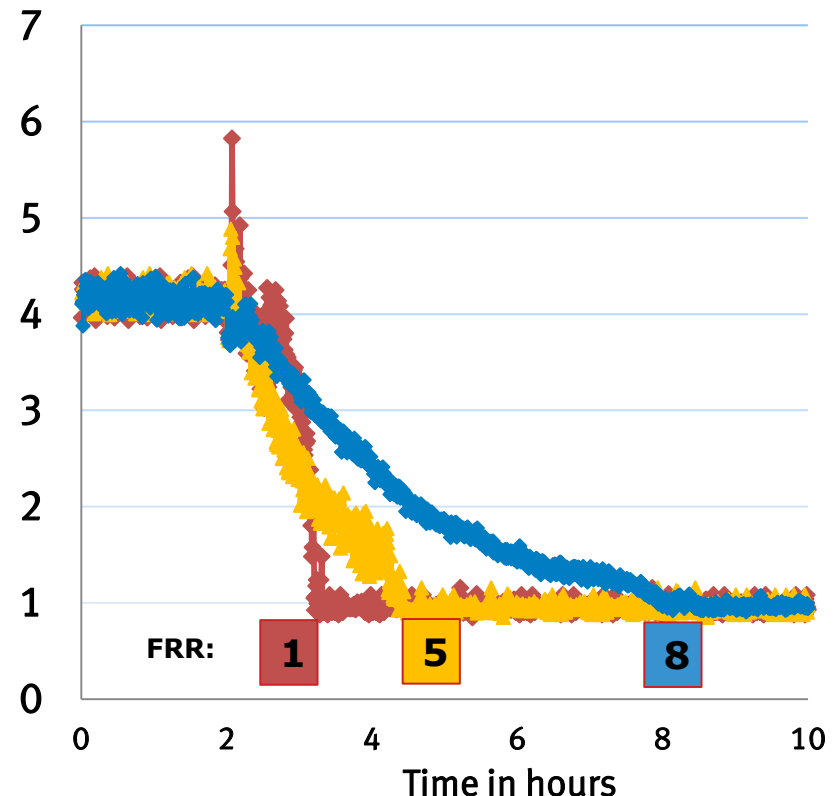
Control Mechanisms

- Storage Group Prioritization – useful when multiple SG's are associated to the same policy and competing for the same resources



High priority is moved to FLASH first

Data Movement Speed



FRR:

1

5

8

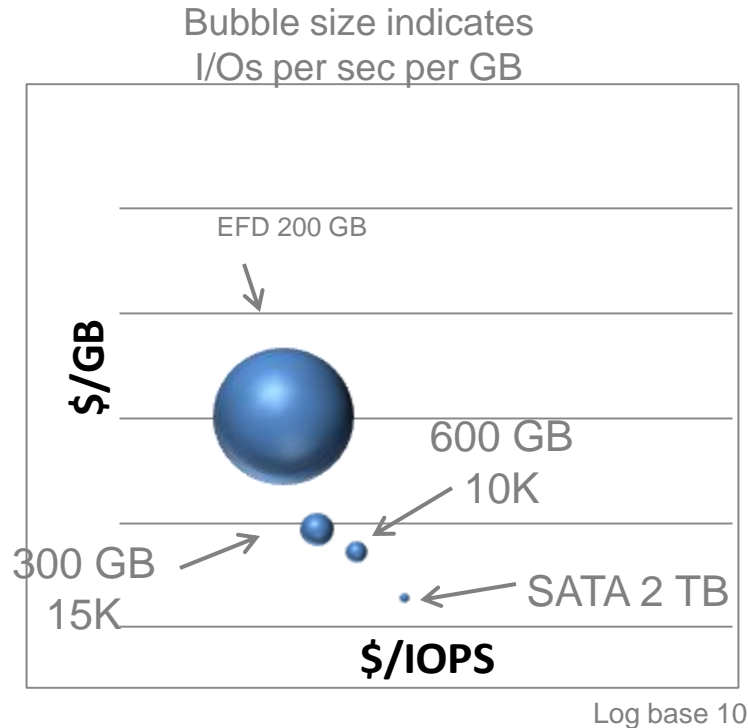
FRR = Fast Relocation Rate
1 is the fastest – 5 is the Default

Configuration Best Practices

Avoiding road construction and traffic jams



Choose the Right Disk Type for the FC/SAS Tier

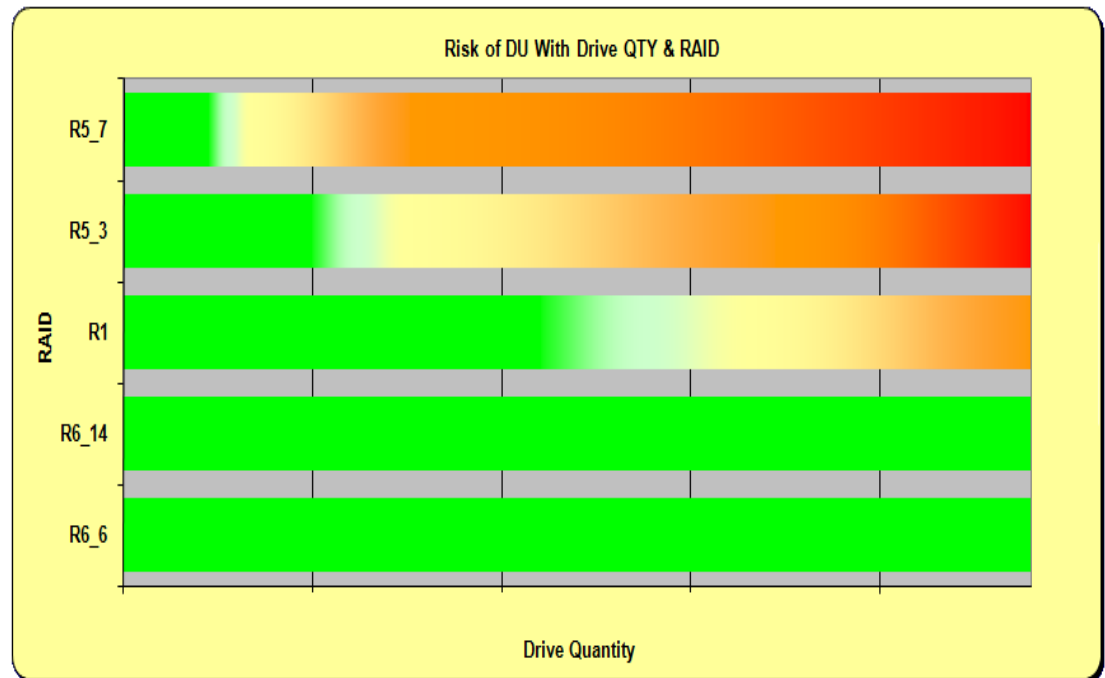


The number of IOPS per GB on a 300 GB 15K drive is ~3X higher than in a 600 GB 10K

- As a general rule of thumb consider smaller and faster FC/SAS Disks in the mid-Tier because of the expected I/O density
- Consider \$/IOPS vs. \$/GB
- But note that we cannot be prescriptive...
 - Best recommendation still is to do the proper analysis to find the drive type that matches the workload
 - Ex: larger drive sizes can be a good match for light workloads

Thin Pool Size

- Overall pool reliability is based on drive quantity and RAID protection
- Reliability can be increased by adding SRDF protection to TDEVs using the pool



*Spinning Drive Example

RAID Protection selection

Protection	Random Write Performance	Resiliency
RAID 1	Best	Better
RAID 5	Better	Good
RAID 6	Good	Best

- Choose **RAID 5 (3+1)** for EFD
 - More usable capacity per drive
- Choose **RAID 1** for FC
 - Consumes less DA resources
 - FC tier is typically less than 70% of total capacity
- Choose **RAID 6 (6+2)** for SATA
 - Greater chance for optimized writes on VMAX 40K

Thin Device Binding

- Bind all thin devices to the FC tier
 - Provides more predictable performance for new allocations
 - Binding to a single tier simplifies management
- Do not bind FAST VP managed thin devices to EFD pools
 - Let FAST VP figure out what data belongs on EFD
- Don't preallocate just because you can
 - Preallocated, but unwritten extents can (and likely will) be demoted
 - Writes to preallocated extents will be on lowest tier

Migration Considerations

Navigation choices



Migration Methods



Host based:

- Logical copy
- Non-disruptive option
- Channel resources must taken into consideration

Array based:

- SRDF/Adcopy
- No host resources
- AutoSwap could be use for a non-disruptive approach

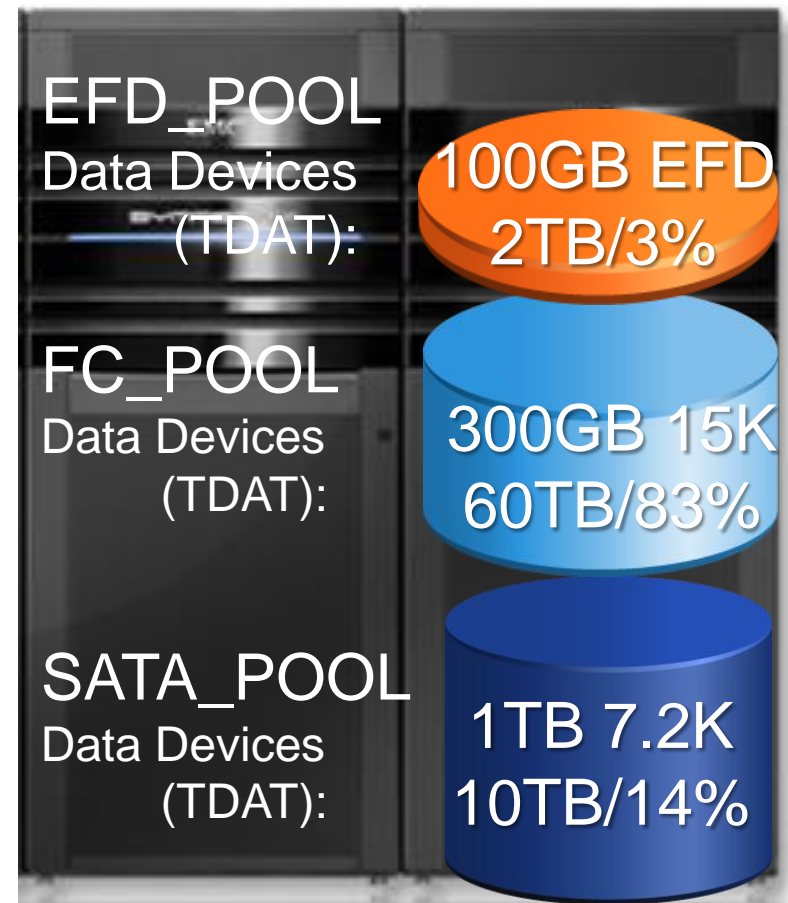
VP Allocation by FAST VP Policy

- When enabled, new allocations can be made in any pool contained within the associated FAST VP policy
 - Alleviates possibility of new allocation failing if chosen pool is full
 - Useful during **migrations** if target tier has insufficient free capacity
- Enable **allocation by policy**
- FAST VP picks the best tier and pool
 - Based on FAST VP metrics and policies
 - Uses policy percentages and pool free space if no metrics

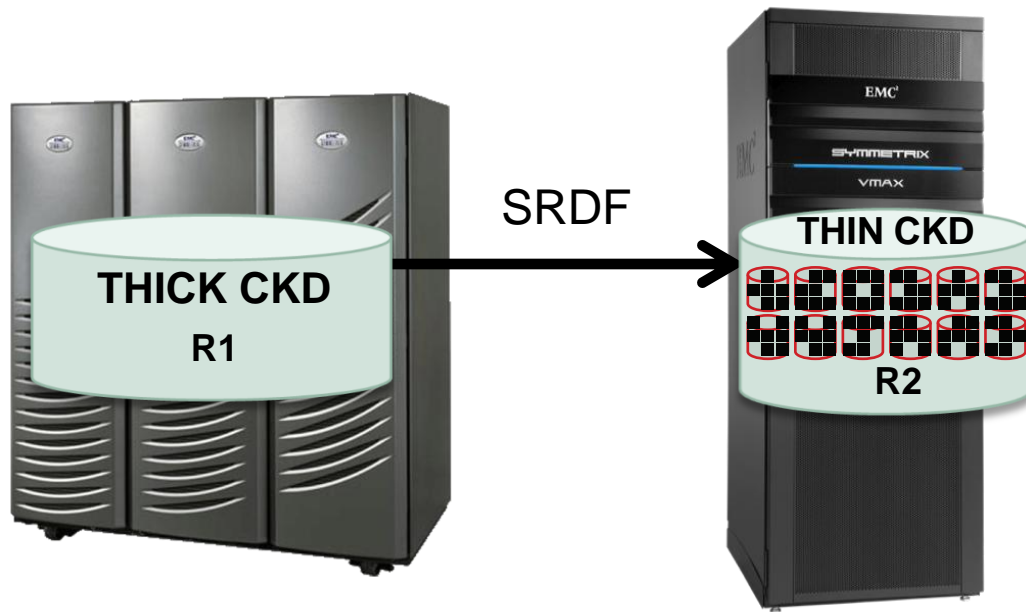
120%Over-subscribed Pool

CU#	SSID	TYPE	CYL	GB/CU	3390B
00	1000	3390-27	32760	6236.16	224
01	1001	3390-27	32760	6236.16	224
↓	↓	↓	↓	↓	↓
0B	100B	3390-27	32760	6236.16	224
0C	100C	3390-9	10017	2042.4	240
0D	100D	3390-3	3339	681.6	240
				71321.76	2944

Thin Devices (TDEV) = ~72TB



SRDF Thick to Thin support - MF



- Support SRDF Thick to Thin (and thin to thick) for MF CKD
 - Refer to SRDF Product Guide for supported Engenuity levels and topologies
- Thin Reclaim Utility support for Thick R1→Thin R2

VLUN VP Mobility and Pinning

- Using VLUN, you can control the location of devices
 - Move all extents of a device to a desired pool
- FAST VP allows pinning all extents of a device in their current locations
 - When pinned, FAST VP does not promote or demote extents of the device
- Leverage these two features to **override FAST VP**
 - For example, move previously active volumes to EFD and pin them in preparation for end-of-quarter processing

Monitoring FAST VP

Unisphere for VMAX Performance



FAST VP Management



The screenshot displays the EMC Unisphere for VMAX web interface. The top navigation bar includes links for Home, System, Storage, Hosts, Data Protection, Performance, and Support. The 'Storage' tab is selected, showing a sidebar with 'Physical Capacity' and 'Alerts' for storage group 000195700398. The main content area, titled '000195700398 > Storage', contains a grid of management options: Storage Groups, FAST, Thin Pools, Tiers, Volumes, Storage Templates, External Storage, and Disk Groups. A red rectangular box highlights the 'Storage Groups', 'FAST', 'Thin Pools', and 'Tiers' options. The text 'You can do everything using SYMCLI, too. But the screenshots are not as pretty!' is overlaid in red across the bottom half of the interface.

EMC Unisphere for VMAX

000195700398 > Home

000195700398 > Storage

Storage Groups
View and manage storage groups. Includes expand, delete and set FAST associations.

FAST
View FAST Dashboard. Includes create FAST policies and place storage groups under FAST control.

Thin Pools
View and manage thin pools. Includes create, bind, write balancing and related objects.

Tiers
View and manage storage tiers. Includes create, delete, tier usage and related objects.

Volumes
View and manage volumes. Includes volume create, convert, set attributes and volume distribution.

Storage Templates
View and manage storage templates. Includes create, import and export.

External Storage
View and manage Federated Tiered Storage (FTS).

Disk Groups
View and manage disk groups. Includes usage, related

**You can do everything using SYMCLI, too.
But the screenshots are not as pretty!**

Unisphere for VMAX Performance

- Unisphere provides real-time, diagnostic, and historical performance monitoring for FAST VP environments
- Monitor view provides system and user-defined dashboards to collate multiple performance indicators into a single view
- Three pre-configured FAST VP dashboards provide views into FAST VP for all primary storage elements
 - FAST VP By Storage Group
 - FAST VP By Tier
 - FAST VP By Policy

Select viewing
date range

Select
dashboard

Create custom dashboard

Generate PDF report



Maximize
chart size

Summary

- Good planning will make the journey much easier
 - Understand the workload and requirements
 - Perform analysis considering SMS group definitions
- Follow configuration best practices
- Consider migration choices
- Monitor for exceptions and watch trends
- Mainframe FAST VP works!

For more Virtual/Tiered storage information:



- Available on <http://support.emc.com>
FAST VP for EMC Symmetrix VMAX Theory and Best Practices for Planning and Performance
- Other SHARE sessions
 - 13983: VMAX Management: Mainframe Tasks Simplified with Unisphere for VMAX & EzSM – Barry Snoots - Wed. @ 4:30 in Room 305
 - 13584: Less=More with Thin Provisioning and Linux on System z – Gail Riley - Thur. @ 4:30PM in Room 308
- EMC.COM Mainframe Page
 - <http://www.emc.com/storage/mainframe.htm>
- Follow EMC Symmetrix
 -  @EMCSymmetrix
 -  www.facebook.com/Symmetrix



A GPS Guide for Your Journey to Automated Storage Tiering

Thank You!

Tony Negro
EMC Corporation

Tuesday August 13, 2013
Session Number 14125

