Getting Even More (and a lot less) From Storage Virtual Provisioning and Automated Storage Tiering

Tony Negro
EMC Corporation

Tuesday, February 05, 2013 1:30 PM
Session Number 12174
Objectives

At the end of this session, you will better understand:

• Virtual Provisioning and Virtual Pools for Count Key Data (CKD) devices
• Advanced utilities for Virtual Pools and array based compression features for CKD devices
• Fully Automated Storage Tiering for Virtual Pool (FAST VP) theory of operation
• Potential benefits that can be achieved when deploying these technologies
Agenda

- Storage Provisioning Overview
  - Provisioning terminology
  - Thin Provisioning mapping
  - Provisioning challenges
- Storage Tiering Consideration
  - Why use different tiers
  - Data mobility between storage tiers
- Automated Storage Tiering Concepts
  - Fully Automated Storage Tiering for Virtual Pool (FAST VP) components
  - FAST VP Implementation
Standard Provisioning Concept

Host addressable devices

Front End (FICON)
Cache
Back End
Disks

Track mapping tables

Single RAID rank

101D-3390-9
101E-3390-9
Host addressable devices

Single RAID rank

Complete your sessions evaluation online at SHARE.org/SanFranciscoEval
Virtual (thin) Provisioning Concept

- Thin devices
- Track group mapping tables
- Virtual Pool
- Data devices
- Front End (FICON)
- Cache
- Back End
- Disks
## Virtual Provisioning Terminology

<table>
<thead>
<tr>
<th>TERM</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Accessible Device</td>
<td>A ‘device’ that is presented on a FICON channel for host use</td>
</tr>
<tr>
<td>Thin Device (Virtual Device)</td>
<td>A ‘Host accessible device’ that has no storage directly associated with it</td>
</tr>
<tr>
<td>Internal Device</td>
<td>A ‘device’ used for internal function of the array</td>
</tr>
<tr>
<td>Data Device</td>
<td>An ‘Internal Device’ that provides storage capacity to be used by a ‘Thin Device’</td>
</tr>
<tr>
<td>Thin Pool (Virtual Pool)</td>
<td>A collection of ‘Data Devices’ that provide storage capacity for ‘Thin Devices’</td>
</tr>
<tr>
<td>Track Group</td>
<td>The size of the smallest contiguous region of a ‘device’ for which an extent mapping can occur</td>
</tr>
<tr>
<td>Bind</td>
<td>The process by which one or more ‘Thin devices’ are associated to a ‘Thin Pool’</td>
</tr>
<tr>
<td>Thin Device Allocated Capacity</td>
<td>The capacity that has been allocated from the thin pool capacity for the exclusive use of a thin device.</td>
</tr>
<tr>
<td>Thin Device Written Capacity</td>
<td>The capacity on a 'Thin Device' that was written to by a host. In most implementations this is a subset of the 'Thin Device Allocated Capacity'.</td>
</tr>
</tbody>
</table>
Standard Provisioning issues

• Slow process

• Complex physical layout

• Physical disk contention ("hot spots")

• Expanding physical capacity can be a challenge
Virtual Provisioned Environment

- 101D-3390-9
- 101E-3390-9
- Thin devices
- Track group mapping tables
- Virtual Pool
- Data devices
- Front End (FICON)
- Cache
- Back End
- Disks
- REBALANCE
Virtual Provisioned Tiered Environment

- Front End (FICON)
- Cache
- Back End
- EFD
- FC
- SATA

Track group mapping tables

Virtual Pools

Thin devices
## 2008 – Fundamental Storage Media Shift

<table>
<thead>
<tr>
<th>Disk</th>
<th>Cost/GB</th>
<th>IOPS/GB</th>
<th>Response Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 K RPM</td>
<td>1</td>
<td>1</td>
<td>6 ms</td>
</tr>
<tr>
<td>Serial ATA (SATA) 7,200 RPM</td>
<td>1/3</td>
<td>1/6</td>
<td>12 ms</td>
</tr>
<tr>
<td>Enterprise Flash Drive (EFD)</td>
<td>8</td>
<td>30</td>
<td>&lt; 1 ms</td>
</tr>
</tbody>
</table>
Virtual Provisioned Tiered Environment

101D - 3390-9

101E - 3390-9

201D - 3390-9

Front End (FICON)

Cache

Back End

EFD

FC

SATA

Track group mapping tables

Virtual Pools

Thin devices

Thin devices (64 P-I-T copies)
SATA Pool thin device allocation and DSS Backup throughput

Running a “special” DB query for the CIO (against the daily P-I-T copy of PROD DB), and it’s REALLY running S_L_O_W….Help!
Let's have a closer look....
Non-disruptive volume mobility (VLUN)

Thin devices

101D-3390-9

101E-3390-9

201D-3390-9

Track group mapping tables

Virtual Pools

 MOVE with REBIND

Front End (FICON)

Cache

Back End

EFD

FC

SATA
Has Random Workload picked up?
Has Response Time improved?

<table>
<thead>
<tr>
<th>STG_GRP</th>
<th>CPU= 16/15 UIC= 65K PR= 0 V11A DEV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACTV RESP IOSQ -DELAY- PEND DISC CONN %D %D</td>
</tr>
<tr>
<td></td>
<td>RATE TIME TIME CMR DB TIME TIME TIME UT RV</td>
</tr>
<tr>
<td>SYMD7A</td>
<td>16C8 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>SYMD7B</td>
<td>16C9 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>SYMD7C</td>
<td>16CA 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>SYMD7D</td>
<td>16CB 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>SYMD7E</td>
<td>16CC 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>SYMD7F</td>
<td>16CD 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>SYMD80</td>
<td>16CE 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>SYMD81</td>
<td>16CF 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>DBSLSG</td>
<td>DBS001 16D0 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>DBSLSG</td>
<td>DBS002 16D1 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>DBSLSG</td>
<td>DBS003 16D2 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>DBSLSG</td>
<td>DBS004 16D3 1.0H 0013 0.000 .000 .000 .00 .00 .00 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS005 16D4 1.5H 0013 142.4 6.27 .000 .01 .00 .123 5.77 .374 58 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS006 16D5 1.3H 0013 715.8 5.01 .000 .02 .00 .131 2.02 .168 21 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS007 16D6 1.5H 0013 169.5 5.58 .011 .01 .00 .124 5.01 .332 60 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS008 16D7 1.0H 0013 173.9 6.06 .109 .01 .00 .122 5.47 .361 T 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS009 16D8 1.0H 0013 0.000 .000 .000 .00 .00 .000 .000 .000 .000 0 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS010 16D9 1.0H 0013 19.15 5.20 .000 .01 .00 .118 4.77 .314 10 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS011 16DA 1.3H 0013 183.9 5.60 .103 .01 .00 .121 5.03 .342 79 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS012 16DB 1.5H 0013 142.6 6.48 .265 .01 .00 .121 5.72 .375 58 0</td>
</tr>
<tr>
<td>DBSDSG</td>
<td>DBS013 16DC 1.0H 0013 0.000 .000 .000 .00 .00 .000 .000 .000 .000 0 0</td>
</tr>
</tbody>
</table>
Support for VP Compression

- Provide configuration, management, and reporting on the VP compression state for thin devices
- NOTE: Compression is supported for devices **NOT** bound with PERSIST
- New device level commands:
  - COMPRESS – compress data for thin device(s)
  - DECOMPRESS – decompress data for thin device(s)
- New parameters on the POOLATTR command:
  - COMPRESSION(ENABLE) – enable compression for a thin pool
  - COMPRESSION(DISABLE) – disable compression for a thin pool
- Note: Compression is disabled by default and must be enabled by issuing the POOLATTR command with COMPRESSION(ENABLE) parameter before thin devices bound to that pool can be compressed
- Compression is a background task that precludes other background tasks from running (such as reclaim or allocation)
Recommendations for Compression

• Compression should be used against very idle data - best not to run medium or greater workloads against it.
  – Can impact the performance

• Auto-compression can only be obtained through use of FAST.

• With the exception of perhaps an occasional backup to tape, it is best to fully decompress data before accessing it.

• Suggested use of manual compression (not FAST) for:
  – archiving old user accounts
  – decompress-use-recompress end of quarter activity
  – use it for low cost/low performance active data
Virtual Provisioning Benefits

- Enables efficient utilization of available resources
  - Virtual Pools
    - Wide striping distributes I/O across spindles and back end processors
    - Advanced utilities allow greater flexibility
  - Provides flexibility when deploying multiple tiers
  - Opportunity for ‘Over provisioning’
    - Provision more space than actually exists
    - Consume space as required
  - Basis for FAST VP
    - Active performance management at a sub-volume, sub dataset level
Automation at all Layers

SMS & HSM

FAST VP
Basis for FAST

- With information growth trends, all Fibre Channel (FC) configurations will:
  - Cost too much
  - Consume too much energy
  - Take up too much space
- FAST 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
  - **Persistence**: If an address range is active (or inactive), it remains so for a while – the longer the duration, the better

80% of IO’s on 20% of capacity
• Symmetrix Tier – a shared storage resource with common technologies (Virtual Pools)
• FAST Policy – manage Symmetrix Tiers to achieve service levels for one or more Storage Groups
• FAST Storage Group – logical grouping of thin devices for common management
FAST VP Time Windows

• **Performance time window** defines when statistics are collected and decayed

• **Workload Analysis Period**
  • Affects decay rates
  • Affects time to respond to changes
  • Default of 7 days (168 hours)

• **Data movement time window** defines when FAST VP is allowed to move data
  • Windows can be customized, but recommendation for initial implementation is 24x7
FAST VP Implementation

- Performance data collected by the system
- **Intelligent Tiering** algorithm generates movement requests based on performance data
- **Allocation Compliance** algorithm generates movement requests based on capacity utilization
- Algorithms continuously assess I/O statistics and capacity use, and make decisions for promotion and demotion
FAST VP – Score Analysis

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

- Start from left, and fill in tiers to determine promotion thresholds

- Allocated Storage

- Prioritized Short Term Score Density

- Low Score Density to fill SATA tier

- Medium Score Density to fill FC tier

- High Score Density to fill EFD tier
Data Movement Granularity Trade-offs

- Larger granularity
  - Uses EFD ineffectively

- Smaller granularity
  - Uses EFD effectively
  - Requires more system resources to maintain statistics

- There is a sweet spot that maximizes the benefits through better use of EFD and reasonable system resource use
Remote Replication Awareness for FAST VP

- Align FAST VP performance between R1 and R2 sites
- Performance metrics are shared from the R1 to the R2 site
- In case there is a swap where production is run from the R2 site, the performance on the R2 site will closely match the R1 performance
Policy-based allocation

- Allocations can come from any pool contained within the tiers of the associated FAST VP policy
  - FAST VP will attempt to allocate from most appropriate tier
    - Based on FAST VP performance metrics and policies
  - If pool is full, alternative pool will be chosen for allocation
    - If array capacity is not oversubscribed new writes should not fail
- If performance metrics exist for the region of the thin device, allocation request will attempt to place new data on the appropriate tier
- If no performance metrics exist request will go to the “bound” pool
- If bound pool is full, allocation request will follow the FAST VP policy percentages
Policy-based allocation – Enabled

Performance Metrics Not Available
Summary

• FAST VP is a policy-based system that promotes and demotes data at the sub-volume, and more importantly, *sub-dataset* which makes it responsive to the workload and efficient in its use of control unit resources.

• FAST VP introduces active performance management, a revolutionary step forward in storage management.

• FAST VP delivers all these benefits without using any host resources.
For more Virtual/Tiered storage information:

- Other SHARE sessions
  - 12945: DB2 for z/OS With EMC Storage Tiering: FAST VP – Wed. @ 8AM in Golden Gate 8
  - 13154: EMC Disk Tiering Technology Review – Wed. @ 12:15 in Golden Gate 2
  - 12708: What’s New With EMC Symmetrix VMAX and Enginuity? – Wed. @ 4:30PM in Golden Gate 7
  - 12317: Less=More with Thin Provisioning and Linux on System z – Thur. @ 3:00PM in Franciscan D

- EMC.COM Mainframe Page
  - [http://www.emc.com/storage/mainframe.htm](http://www.emc.com/storage/mainframe.htm)

- Follow EMC Symmetrix
  - @EMCSymmetrix
  - [www.facebook.com/Symmetrix](http://www.facebook.com/Symmetrix)
Getting Even More (and a lot less) From Storage Virtual Provisioning and Automated Storage Tiering

Tony Negro
EMC Corporation

Tuesday, February 05, 2013 1:30 PM
Session Number 12174

Thank You