A day in the life of a Dell Compellent Page: How Dynamic Capacity, Data Instant Replay and Data Progression work together

Troy Presler, Product Management
Mike Klemm, Engineering Technologist
Compellent Fluid Data Architecture

**Virtualization**
- Simplify management
- Ideal for virtual servers
- Independence from drive type, interface type and RAID

**Intelligence**
- Block-level intelligence
- Manages data inside the volume
- Tracks frequency of access
- Enables unique feature/benefits

**Automation**
- Sophisticated data movement engine
- Set policies and let the system run
- Optimized data placement

**Utilization**
- “No waste” when storing data
- Buy fewer disk drives
- Save on energy and floor space

---

<table>
<thead>
<tr>
<th>SSD</th>
<th>FC</th>
<th>SAS</th>
<th>SATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAID 10</td>
<td>RAID 5</td>
<td>RAID 6</td>
<td></td>
</tr>
<tr>
<td>15K</td>
<td>10K</td>
<td>7.2K</td>
<td></td>
</tr>
</tbody>
</table>

| Tier 1 |
| Tier 2 |
| Tier 3 |

**METADATA**
Data about the data.

**Disks not purchased**
Unused allocation
Compellent Fluid Data Architecture

• Page Pool
  – Collection of allocated and unallocated disk blocks
  – Maps pages to volumes
  – Maintains metadata
  – Default page size is 2MB (4,096 blocks)

• Automated, Sophisticated Block Management
  – Change RAID levels by re-pointing pages
  – System manages data mapping across multiple RAID levels
  – Individual files can span multiple drives
  – Supports multiple drive types and RAID levels
  – Re-stripe data when adding drives

• Page Pool grows/shrinks as needed
  – Self defragmentation and tuning
Data progression

- Data Progression’s data mover runs once per day
  - Default start time is 7 p.m.
  - Data is moved per page
  - Historical Replay pages eligible to move to lowest tier immediately
  - 12 days down, 3 up
  - Data Progression will run against multiple volumes at the same time

- RAID Restripe occurs:
  - When adding additional drives
  - When RAID extents score low
Storage Profiles

Select the classes of storage to be used for Dual-Redundant writable and replay data.

### Tier 1 Storage

<table>
<thead>
<tr>
<th>RAID Level</th>
<th>Writable Data</th>
<th>Replay Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAID 10-DM</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>RAID 6-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID 6-10</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

### Tier 2 Storage

<table>
<thead>
<tr>
<th>RAID Level</th>
<th>Writable Data</th>
<th>Replay Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAID 10-DM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID 6-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID 6-10</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

### Tier 3 Storage

<table>
<thead>
<tr>
<th>RAID Level</th>
<th>Writable Data</th>
<th>Replay Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAID 10-DM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID 6-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID 6-10</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Available RAID levels

- The RAID levels the page pool will create within a tier is defined by the tier’s redundancy level
- Each tier has its own redundancy level, can be different for each tier
- Single redundancy configures the tier for RAID10, RAID 5/5, and RAID 5/9
- Dual redundancy configures the tier for RAID10DM, RAID 6/6, and RAID 6/10
- By default, any tier with a drive size of 900GB or larger will be configured for dual redundancy (new systems)
- Redundancy level can be modified on the fly
Two types of pages: Accessible and Historical (Replay)

- Accessible pages are pages that can be read or written by a server at the current time
- Historical pages are read-only pages (set by Data Instant Replay)

Data Progression uses the accessibility to determine the class of storage a page should use
Levels of optimization

- Tiers
  - Tier 1, Tier 2, Tier 3
- Disk Zones
  - Fast and Standard
- RAID Levels
  - Single Redundancy
    - RAID 10, RAID 5-5 (4+1), RAID 5-9 (8+1)
  - Dual Redundancy
    - RAID10DM, RAID6-6 (4+2), RAID 6-10 (8+2)
- Page Sizes
  - 512KB, 2MB, 4MB
Page lifecycle

• How it all works together
  – Storage system has three tiers of disk
  – Volumes are configured to use Recommended Storage Profile
  – Tier 1 is single redundancy; Tier 3 is dual redundancy
  – Replays are taken multiple times a day and are retained for 3 days
  – Volume was created today
  – Volume size is 1TB
### Page Lifecycle

- **8AM Monday**: Replay is taken. Reads and writes continue, but only changes are written.
- **10AM Monday**: Replay is taken. Reads and writes continue, but only changes are written.

#### Data Progression

<table>
<thead>
<tr>
<th>Time</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>8AM</td>
<td>Replay, Reads and writes continue, only changes written</td>
</tr>
<tr>
<td>10AM</td>
<td>Replay, Reads and writes continue, only changes written</td>
</tr>
</tbody>
</table>

#### RAID Tiers

- **Tier 1**: RAID 10 (Active Writes)
- **Tier 2**: RAID 5
- **Tier 3**: RAID 6 (MLC)

#### Historical Reads

- Tier 1 RAID 10
- Tier 2 RAID 5
- Tier 3 RAID 6

---

**On Demand Data Progression**

- Data is written throughout the day.
- **8AM Monday**: Replay is taken.
  - Reads and writes continue, only changes are written.
- **10AM Monday**: Replay is taken.
  - Reads and writes continue, only changes are written.

---

*Optimize your enterprise*

Melbourne | 19 November 2013
Pages: Accessible, Historical

Accessible, Recently Accessed
- These are the active pages the volume is using the most

Accessible, Non-recently Accessed
- Read/Write pages that have not been recently used

Historical, Recently Accessed
- Read-only pages the volume is using the most
- Think Replays

Historical, Non-recently Accessed
- Read-only pages that have not been recently used

Historical, Non-Accessible
- Read-only data pages that are not currently accessible by a volume
- That is, they have been aged due to a newer version of the page
- System maintains these pages for recovery purposes and are placed on the lowest cost storage possible
Page lifecycle

- 12PM Mon Replay is taken
- On Demand Data Progression runs
- At some point 8AM Mon Replay expires
  - 10AM Mon Replay remains
  - Page C released back page pool

### Diagram

- **12PM Wed**
  - Active Writes
  - Historical Reads

- **10AM Mon**
  - TIER1 RAID10
  - OR
  - TIER2 RAID5
  - MLC
  - Historical Reads

- **8AM Mon**
  - DATA

### Table

<table>
<thead>
<tr>
<th>Time</th>
<th>Tier</th>
<th>RAID</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>12PM</td>
<td>Tier 1</td>
<td>RAID10</td>
<td>Active Writes</td>
</tr>
<tr>
<td>10AM</td>
<td>Tier 1</td>
<td>RAID10</td>
<td>Historical Reads</td>
</tr>
<tr>
<td>8AM</td>
<td>Tier 3</td>
<td>RAID6</td>
<td>Historical Reads</td>
</tr>
</tbody>
</table>

### Workspace

- **C**
  - Write C3...
  - Read C2
- **E**
  - Write E2...
  - Read E1
- **Δ Changes**
  - C1
  - C2
  - E1
- **Replay**
  - C2
  - E1

---

12
Page lifecycle
Recovering a volume

- Recover a Replay
  - This becomes a new branch
  - The new branch shares read-only blocks
Page lifecycle Recovering a volume

12PM Mon
DATA

10AM Mon
DATA

Volume 1
READ A
READ C2
WRITE C3
READ E1
WRITE E2

Volume 2
READ A
READ C2
WRITE C4
READ E1
WRITE E3

Time 3

Recovery View
Page lifecycle
Unmap/Space Recovery

- Unmap Pages
  - Unmap Request to Active Replay for pages D & E
    - Return E2 to Page pool
    - Mark blocks as unmapped
- Expire 12PM Mon Replay
- C2 & E1 released back to Page pool
- 10AM Mon Replay Remains
  - Coalesce Replay

12PM Mon
Unmap

10AM Mon
DATA
Unmap

TIER1 RAID10
Active Writes

TIER1 RAID10
Historical Reads

TIER1 RAID5
Historical Reads

MLC TIER2 RAID5
Historical Reads

TIER3 RAID6
Historical Reads
Storage Center 6.4 data progression enhancements

“Optimized for Flash Environments”

- On Demand Data Progression
  - Optimized for all flash environments
    - Leverages Read Intensive SSD tier
    - Also applies to spinning disk
  - Runs after replays are taken
    - Frozen blocks rapidly converted to RAID 5
    - RAID 5 blocks quickly progressed to Read Intensive (RI) MLC drives in flash environments

- Space Management Replays
  - Monitor and manage Tier 1 RAID 10 space
  - Initiate replay when less than 5% space free
  - On demand data progression runs after replay
Space management replay - Page lifecycle

- Data is written throughout the day
- Space Management Replay is taken
- Reads and writes continue
  - Only changes are written
- On Demand Data Progression runs
- 6PM Tue Replay is taken
- On Demand Data Progression runs
- Space Management Replay Coalesce

6PM Tue
DATA

TIER1 RAID10
Active
Writes

TIER1 RAID10
Historical
Reads

TIER1 RAID5
MLC
OR
TIER2 RAID5
Historical
Reads

TIER3 RAID6
Historical
Reads

Δ Changes
Replay

Δ Changes
Space Management
Replay

C1
Best Practices

• Dell Compellent recommends following our best practices to maximize the advantage these features
• Replay best practice
  – All volumes should have at least one replay scheduled per day
  – Exception is log, swap, and pagefile volumes – no need to progress this data
• Volume best practices
  – Configure all volumes for the Recommended Storage Profile
  – Exception is log, swap, and pagefile volumes – suggest the High Priority (Tier1) Profile
• Resources:
  – Presentation is available for registered attendees
Thank you!
Notices & Disclaimers

Copyright © 2013 by Dell, Inc.

No part of this document may be reproduced or transmitted in any form without the written permission from Dell, Inc.

This document could include technical inaccuracies or typographical errors. Dell may make improvements or changes in the product(s) or program(s) described herein at any time without notice. Any statements regarding Dell’s future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

References in this document to Dell products, programs, or services does not imply that Dell intends to make such products, programs or services available in all countries in which Dell operates or does business. Any reference to an Dell Program Product in this document is not intended to state or imply that only that program product may be used. Any functionality equivalent program, that does not infringe Dell’s intellectual property rights, may be used.

The information provided in this document is distributed “AS IS” without any warranty, either expressed or implied. Dell EXPRESSLY DISCLAIMS any warranties of merchantability, fitness for a particular purpose OR INFRINGEMENT. Dell shall have no responsibility to update this information.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any Dell patents or copyrights.

Dell, Inc.
300 Innovative Way
Nashua, NH 03063 USA