Better resilver – progress 1 yr on

SAŠO KISELKOV | kernel engineer
Better resilver
Better resilver

On initial writing

Logical block order
- #1
- #2
- #3
- #4
- #5
- #6
- #7

Physical block order
- #1
- #2
- #3
- #4
- #5
- #6
- #7

Resilver order

Resilver order

After lots of rewriting

Logical block order
- #1
- #2
- #3
- #4
- #5
- #6
- #7

Physical block order
- #3
- #2
- #4
- #1
- #7
- #5
- #6

Resilver order

Resilver order
Better resilver
Better resilver

Further improvements

- Shipped in February
- New default algorithm on NexentaStor 5.0.3 and newer
- No disk format change
- Involved Matt & George in fine-tuning the design
- Tom Caputi from Datto improved resilver prefetcher (ZoL)
Better resilver

Performance

Promises, promises...

Scrub time (lower is better)

998 GB data (71% cap), 1x5 raidz2 (300GB HDDs)

- stock 34:20
- improved (queuesz=262M) 02:05 (16.5x)
Better resilver

Performance

raw data size: 21.1 TiB (12% full)

recordsize: 128k

pool geometry: 12 x 8-drive raidz2, 2x slog (200GB SAS SSD)

main pool devices: SEAGATE ST2000NM0023 1.82TB

<table>
<thead>
<tr>
<th>ALGO</th>
<th>TEST</th>
<th>TIME (hh:mm)</th>
<th>RATE (MB/s)</th>
<th>DELTA</th>
<th>CACHESZ (MB)</th>
<th>GAPSZ (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>improved</td>
<td>scrub</td>
<td>10h44m</td>
<td>573</td>
<td>3.1x</td>
<td>26121</td>
<td>2</td>
</tr>
<tr>
<td>stock</td>
<td>scrub</td>
<td>32h45m</td>
<td>188</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>
Better resilver

Performance

raw data size: 26.2 TiB (pool 36% full)

recordsize: 16k

pool geometry: 2 x 5-drive raidz2

devices: 10x HGST 8TB 7k2 3.5" SAS (HUH728080AL5204)

<table>
<thead>
<tr>
<th>KERNEL</th>
<th>TEST</th>
<th>TIME (hh:mm)</th>
<th>RATE (MB/s)</th>
<th>DELTA</th>
<th>CACHESZ (MB)</th>
<th>GAPSZ (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>stock</td>
<td>scrub</td>
<td>49h16m</td>
<td>154.9</td>
<td>⎯</td>
<td>⎯</td>
<td>⎯</td>
</tr>
<tr>
<td>improved</td>
<td>scrub</td>
<td>14h40m</td>
<td>520</td>
<td>3.4x</td>
<td>1309</td>
<td>2</td>
</tr>
<tr>
<td>improved</td>
<td>scrub</td>
<td>13h49m</td>
<td>552</td>
<td>3.6x</td>
<td>6548</td>
<td>2</td>
</tr>
</tbody>
</table>
Better resilver

Performance

Customer boxes (X speedup from old algo)
Better resilver

Performance

- Effects of randomization

![Graph showing performance comparison between new and old scrub methods.](image-url)
Better resilver

Performance

- Effects of randomization

![Graph showing performance comparison between new and old resilver methods.](image)