Modernization of Storage Scale: Dynamic Pagepool

Storage Scale UK User Group Meeting 2023
London, UK – June 27-28, 2023

Christof Schmitt <christof.schmitt@us.ibm.com>

Disclaimer

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

IBM reserves the right to change product specifications and offerings at any time without notice. This publication could include technical inaccuracies or typographical errors. References herein to IBM products and services do not imply that IBM intends to make them available in all countries.

Storage Scale pagepool

- Used to cache user data and filesystem metadata in memory.
- The size of the pagepool limits the amount of data or metadata that can be cached without requiring I/O to disk.
- The larger the pagepool, the less expected amount of I/O calls to disk.

Static pagepool

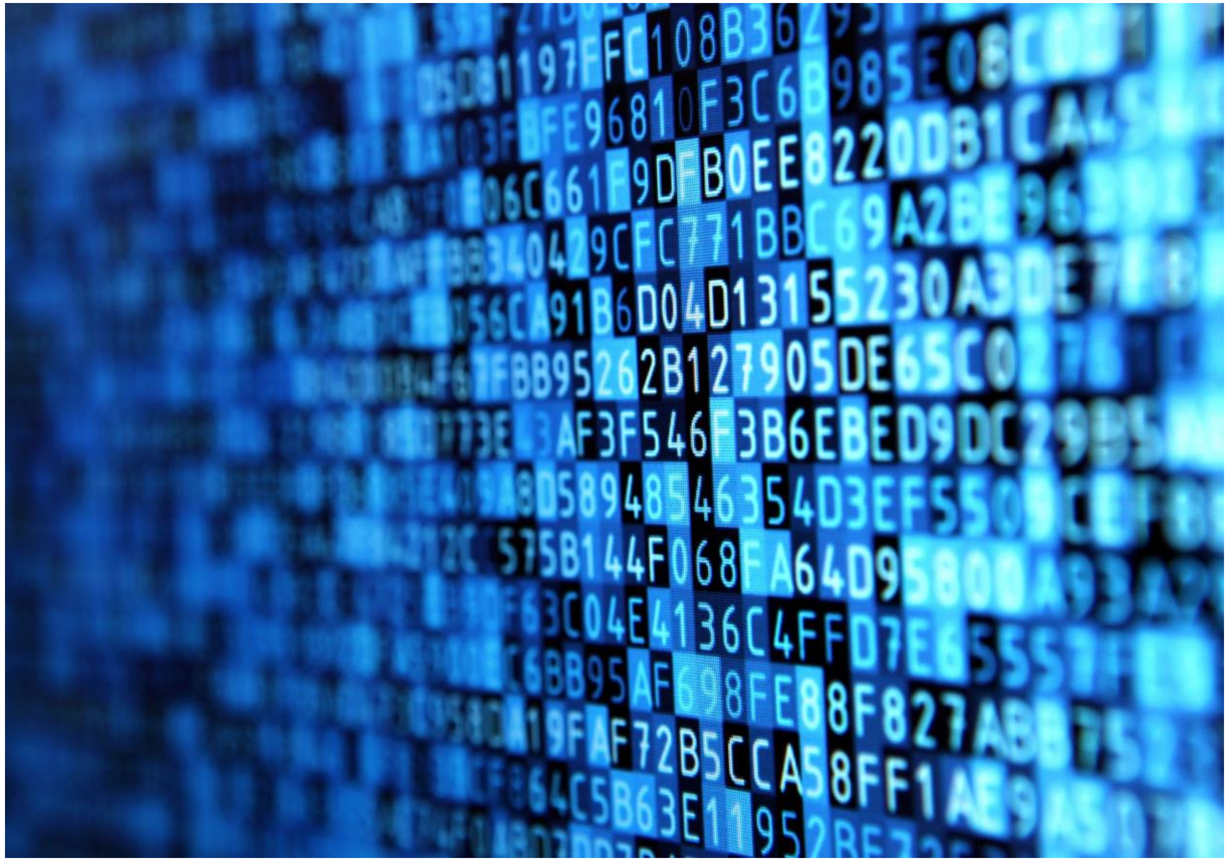
- Only mode up to Scale 5.1.6
- Default mode in 5.1.7 and 5.1.8
- Set to static size with `mmchconfig pagepool=…`
- Fixed amount of memory reserved on startup, not available to applications.
- Setting too small: Impacts performance due to cache misses
- Setting too large: Limits available application memory
- Does not handle change in demands during runtime.

Dynamic Pagepool

- New way of managing size of Storage Scale pagepool
- Available as Tech Preview in Scale 5.1.7 and 5.1.8
- When enabled, pagepool size is adjusted dynamically between minimum and maximum boundaries
- No further configuration necessary for most cases, but boundaries can be adjusted for special cases.
- Dynamic Pagepool must not be used in production environments during the Technical Preview; it is intended for test environments only.

High-level behavior of Dynamic Pagepool

- Spectrum Scale registers a callback with the Linux kernel to receive memory pressure notifications.
- Attempt to reduce pagepool size upon receiving memory pressure notification
- Never shrink below configured minimum size
- Upon repeated fetch of the same data into pagepool, attempt to grow pagepool.
- Never grow beyond configured maximum size



Storage Scale

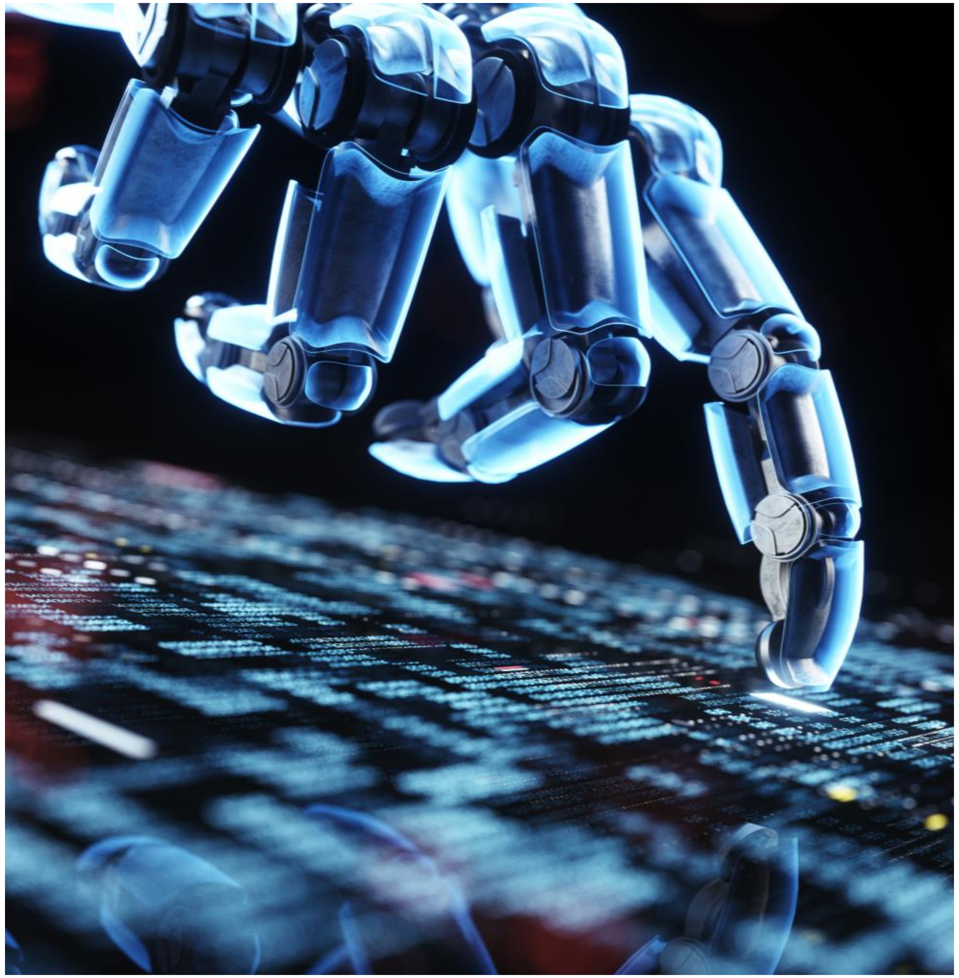
Dynamic Pagepool Simple Configuration

```
mmchconfig dynamicPagepoolEnabled=yes -N node1

mmchconfig pagepool=default -N node1

mmshutdown -N node1

mmstartup -N node1
```



Monitoring size of Dynamic Pagepool

A new mmdiag command reports the size of the dynamic pagepool in 5.1.7 and 5.1.8. The plan is to provide zimon monitoring for the pagepool size when the dynamic pagepool becomes generally available.

```
# mmdiag --pagepool
=== mmdiag: pagepool ===
Dynamic pagepool: enabled
Minimum pagepool size: 407022592 Bytes (397483 KiB, 388 MiB, 0 GiB)
Current pagepool size: 3221225472 Bytes (3145728 KiB, 3072 MiB, 3 GiB)
Maximum pagepool size: 6105326592 Bytes (5962233 KiB, 5822 MiB, 5 GiB)
Physical memory  size: 8140435456 Bytes (7949644 KiB, 7763 MiB, 7 GiB)
```

```
# mmdiag --pagepool -Y
mmdiag:pagepool:HEADER:version:reserved:reserved:dynamicPagepool:minimumSize:c
urrentSize:maximumSize:physicalMemorySize:
mmdiag:pagepool:0:1:::1:407022592:3221225472:6105326592:8140435456:
```

Dynamic Pagepool config parameters

Config parameter	Allowed values	default	Description
`dynamicPagepoolEnabled`	yes/no	no	Enable dynamic pagepool vs. static pagepool
`pagepoolMinPhysMemPct`	1 - 50	5	Minimum size of dynamic pagepool as percentage of physical memory.
`pagepoolMaxPhysMemPct`	10 - 90	75	Maximum size of dynamic pagepool as percentage of physical memory.

Storage Scale

Dynamic pagepool limitations for Techical Preview

- Linux only
- Dynamic pagepool can only be used on **client nodes**, not on NSD or ECE server nodes. This is not enforced during the Technical Preview. The plan is to enforce this when the feature becomes generally available.
- **fsck** might fail when the dynamic pagepool is not close to the maximum size. The suggestion for the Technical Preview is to limit fsck to nodes without the Dynamic Pagepool enabled. The plan is to implement proper co-existence for fsck and the dynamic pagepool when the feature becomes generally available.
- Dynamic Pagepool Technical Preview is not supported together with **RDMA**. Ensure that RDMA is not enabled for the nodes with the dynamic pagepool enabled. The plan is to support RDMA also on nodes with the dynamic pagepool when the feature becomes generally available.

How to get involved

- Dynamic pagepool available as Technical Preview in 5.1.7 and 5.1.8
- 5.1.8 contains one important fix
- https://supportcontent.ibm.com/support/pages/node/6956570
- Formally sign-up for Technical Preview at scale@us.ibm.com to get in contact with development and provide feedback
- Interested to hear about workloads with varying memory requirements

Thank you for using IBM Storage Scale!