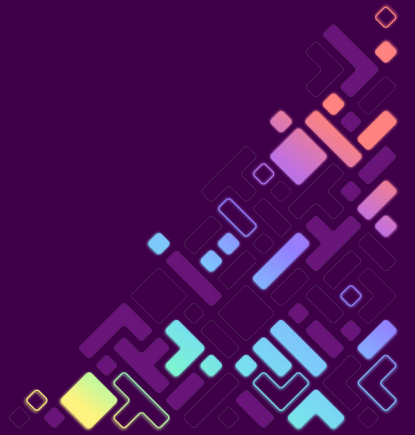
LINUXCON @ OPEN SOURCE SUMMIT EUROPE

THE LINUX FOUNDATION

DAMON Recipes:

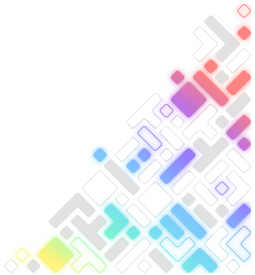
Ways to Save Memory Using a Linux Kernel Subsystem in the Real World

SeongJae Park <sj@kernel.org>
Honggyu Kim <honggyu.kim@sk.com>



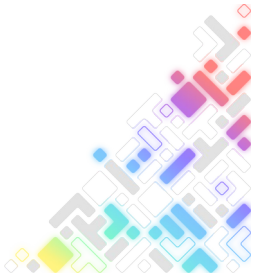
Notices

- The views expressed herein are those of the speakers;
  They do not reflect the views of their employers
- This talk is for DAMON recipes (usages);
  detailed internals of DAMON are out of the focus

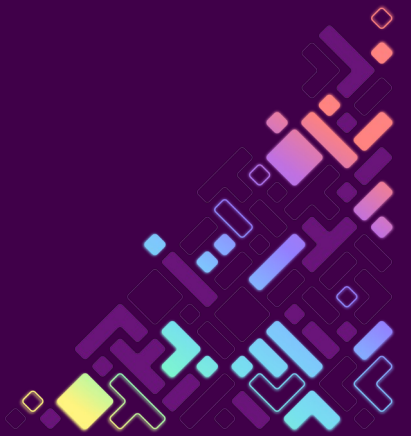


Overview

- DAMON in Nutshell (5 mins)
- DAMON Recipes
  - Profiling and profiling-guided optimizations (5 mins)
  - Memory auto-scaling in AWS Aurora Serverless v2 (7 mins)
  - Tiered memory management in SK hynix HMSDK (15 mins)
- DAMON Community (2 mins)
- Summary (1 mins)
- QnA (5 mins)

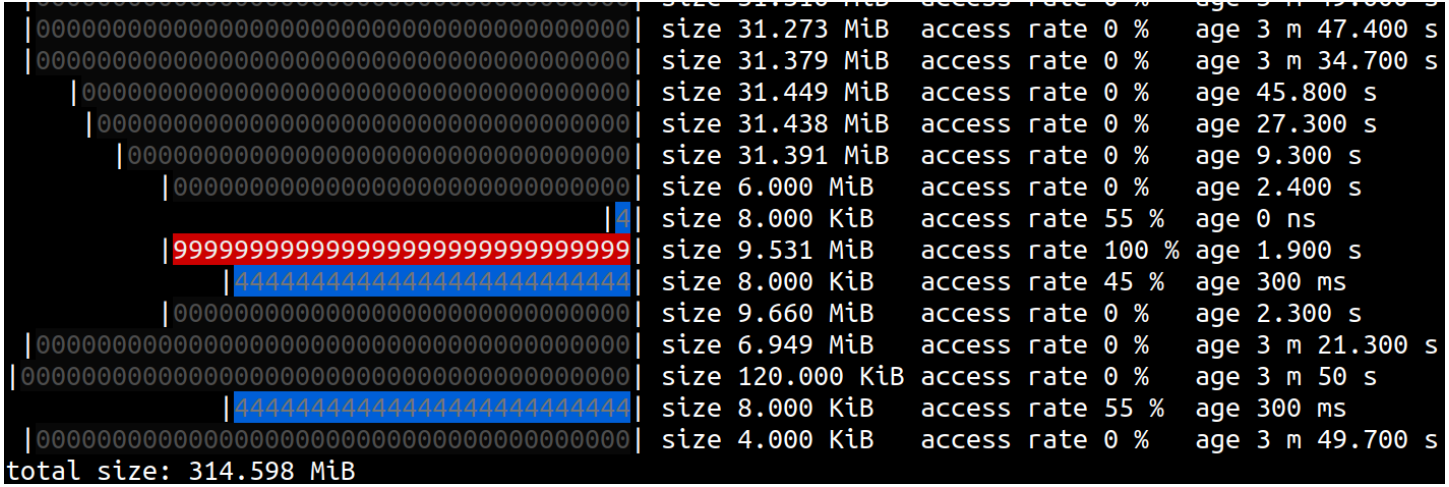


DAMON in Nutshell



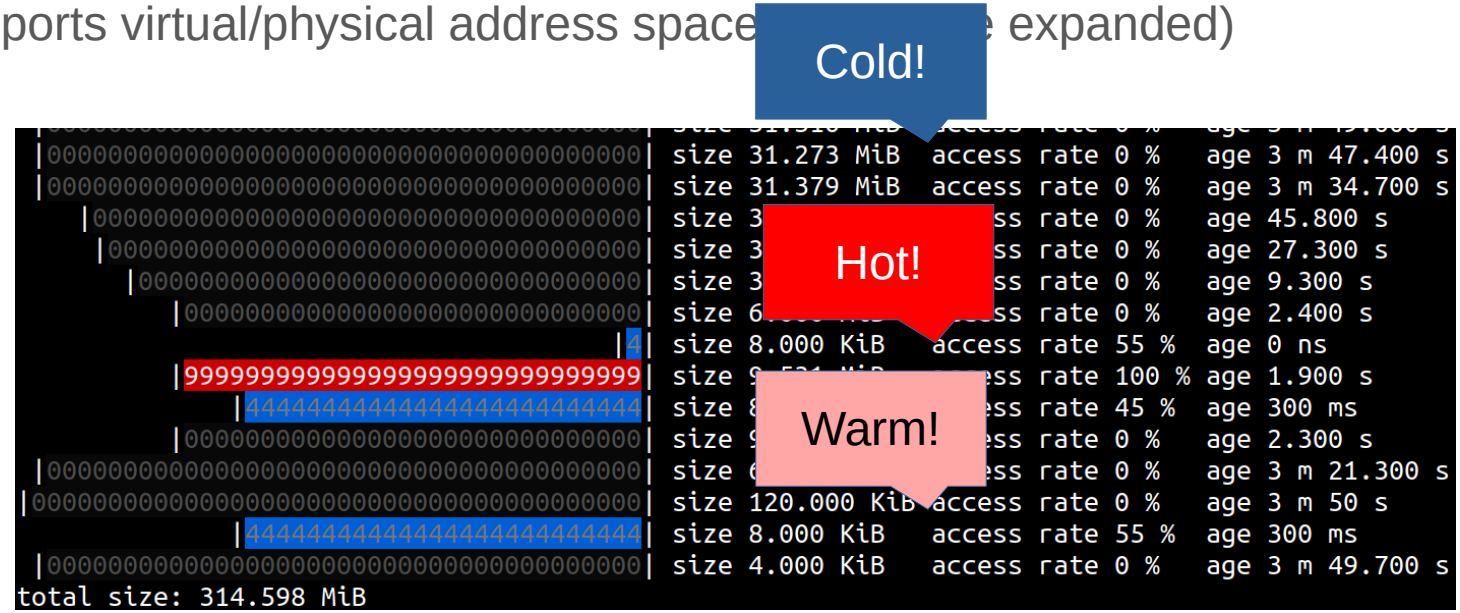
DAMON: Data Access Pattern Snapshot Generator

- Informs *which address range* is *how frequently accessed* for *how long time*
- Supports virtual/physical address spaces (could be expanded)



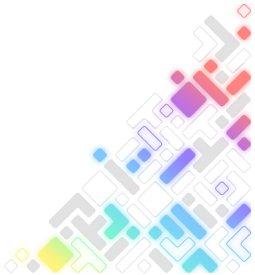
DAMON: Data Access Pattern Snapshot Generator

- Informs *which address range* is *how frequently accessed* for *how long time*
- Supports virtual/physical address space (and expanded)



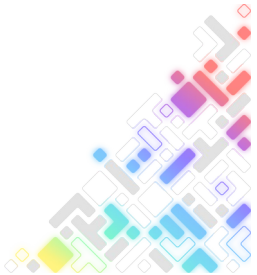
Best-effort Overhead-Accuracy Tradeoff

- Let users control upper-bound monitoring overhead limit
- Under the limit, provides its best accuracy using adaptive mechanisms
- From real world production usages, 3-4% CPU usage is commonly reported
- More details available at the design doc



Extensible Design

- DAMON supports virtual address spaces and the physical address space
- The core logic and address space/access check-specific logic are separated
- Can be extended for different address spaces and access check primitives
- E.g., CPU (AMD IBS, Intel PEBS, …) or devices (e.g., GPU, CXL, …) providing access check primitives for special address spaces (e.g., CPU cache, unified memory, low-tier memory, …)
- More details are available at design doc



Extensible Design

- DAMON supports virtual address spaces and the physical address space
- The core logic and address space/access check-specific logic are separated
- Can be extended for different address spaces and access check primitives
- E.g., CPU (AMD IBS, Intel PEBS, …) or devices (e.g., GPU, CXL, …) providing access check primitives for special address spaces (e.g., CPU cache, unified memory, low-tier memory, …)
- More details are available at design doc

DAMOS: DAMon-based Operation Schemes

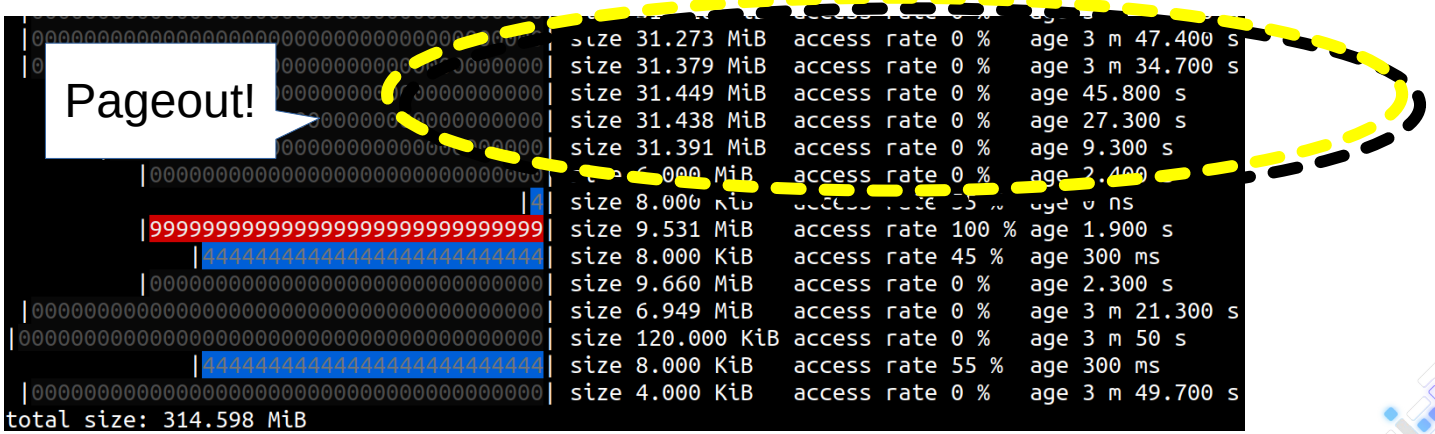
- A feature of DAMON
- Apply memory operation actions to regions of interesting access pattern
- More details are available at design doc

```
# # pageout memory regions that not accessed for >=5 seconds
# damo start --damos_action pageout --damos_access_rate 0% 0% --damos_age 5s max
```

DAMOS: DAMon-based Operation Schemes

- A feature of DAMON
- Apply memory operation actions to regions of interesting access pattern
- More details are available at design doc

```
# # pageout memory regions that not accessed for >=5 seconds
# damo start --damos_action pageout --damos_access_rate 0% 0% --damos_age 5s max
```



Availability

- DAMON is enabled on kernels of major Linux distros
- including Amazon Linux, Android, Debian, Fedora, Oracle Linux
- DAMON user-space tool is packaged for multiple packaging systems including Arch, Debian, Fedora, PyPi, Raspbian

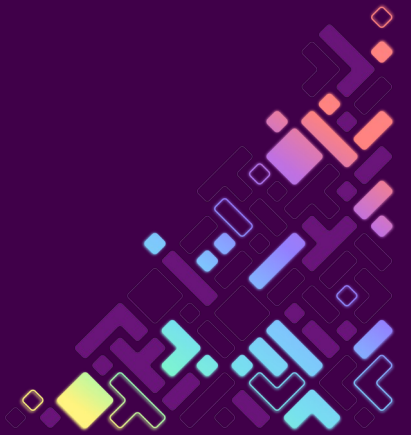
Distribution	DAMON [x]	UTS_RELEASE [x]
Android 12 (5.10) aarch64	y	5.10.218
Android 13 (5.10) aarch64	y	5.10.218
Android 13 (5.15) aarch64	y	5.15.151
Android 14 (5.15) aarch64	y	5.15.158
Android 14 (6.1) aarch64	y	6.1.90
Arch x86_64	y	6.10.9-arch1
CentOS 9 Stream aarch64	y	5.14.0-505.el9.aarch64
CentOS 9 Stream x86_64	y	5.14.0-505.el9.x86_64
CentOS Hyperscale 9 aarch64	y	6.10.9-0.hs1.hsx.el9.aarch64
CentOS Hyperscale 9 x86_64	y	6.10.9-0.hs1.hsx.el9.x86_64

Packaging status	
AOSC	2.4.7
AUR	2.4.0
Debian 13	2.4.9
Debian Unstable	2.4.9
Devuan Unstable	2.4.9
EPEL 9	2.4.8
EPEL 10	2.4.9
Fedora 37	1.9.8
Fedora 38	2.3.3
Fedora 39	2.4.8
Fedora 40	2.4.8
Fedora Rawhide	2.4.9
Kali Linux Rolling	2.4.9
PureOS landing	2.4.2
PyPI	2.4.9
Raspbian Testing	2.4.9
Ubuntu 23.10	1.8.7
Ubuntu 24.04	2.2.4
Ubuntu 24.10	2.4.6

That’s It About DAMON Itself

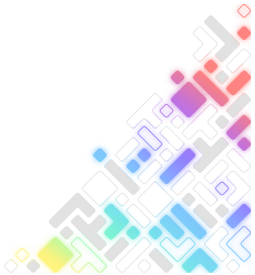
- Internal mechanism of DAMON is out of the scope of this talk
- This talk is for how {others are using,you can use} DAMON for fun and profit
- For more details of the internal mechanism, refer to the design doc

Three DAMON Recipes

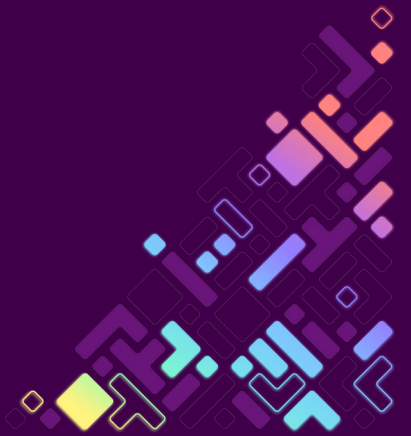


Introducing (Only) Three DAMON Recipes

- No one knows entire and optimum DAMON usages
- Following three recipes are only speakers' best knowledge
- We are waiting to learn more recipes from you



Accesses Profiling (-guided Optimization)



DAMON for Profiling (and Optimizations)

- Record, or snapshot data access pattern of workloads using DAMON
- Record/snapshot additional information together
- Visualize the data points (e.g., heatmap, flamegraphs, workingset size)
- Get insights from data for better understanding workloads and systems
- Make profiling-guided optimizations

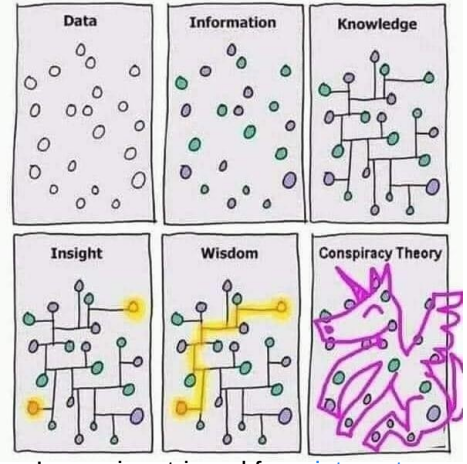
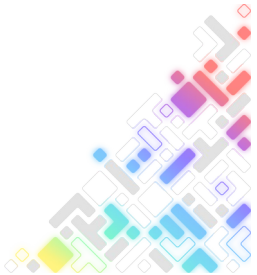


Image is retrieved from internet

General DAMON-based Profiling-Guided Optimization

- Find hot/cold regions and matching data/variables in the code
- Protect/promote hot data using `mlock()` or `madvise(MADV_HUGEPAGES)`
- Evict/demote cold data using `madvise(PAGEOUT,NOHUGEPAGES)`
- Any creative optimizations would be possible
- Hot page protection achieves 2.3x speedup under memory pressure
- More detailed blog article

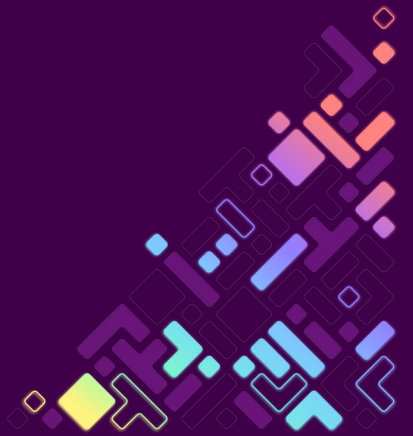


DAMON user-space tool supports for PGO

- DAMON user-space tool, `damo`, is for easy usages of DAMON
- We are adding features for easy DAMON-based PGO to damo
- As of this talk, some profiling parts are added, more works to go
- Quick start: `damo monitor --report_type holistic`

Generated from
https://qr.io

Proactive Reclamation

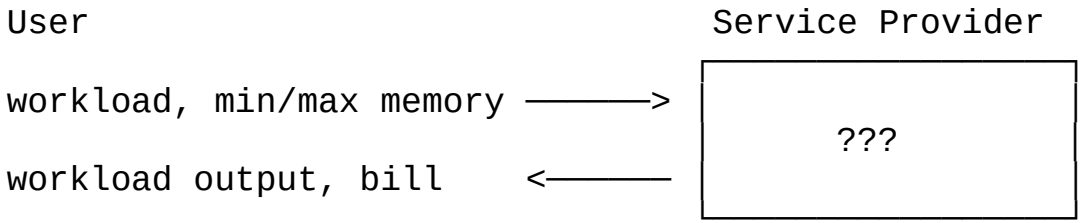


Proactive Reclamation

- Reactive reclamation: Reclaim cold memory when memory pressure happens
- Proactively reclamation: Reclaim cold memory before memory pressure
- Benefit 1: Reduce memory footprint without performance degradation
- Benefit 2: Minimize degradation from direct reclamation
- Known usages: Google, Meta, and Amazon
- Each company uses its own implementation for its usage
- Amazon uses DAMON-based implementation for memory auto-scaling

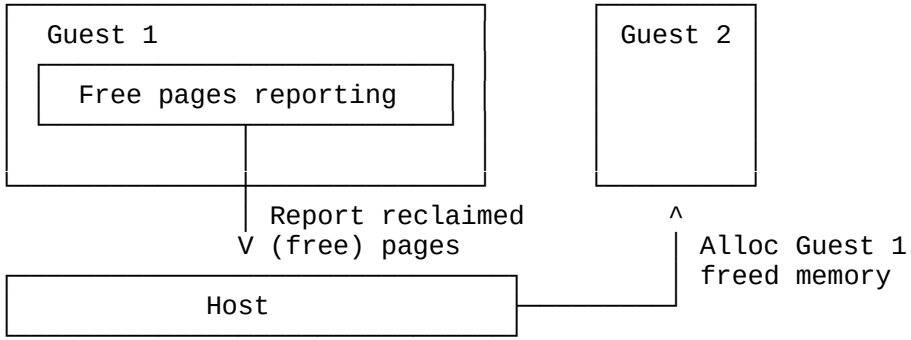
Memory Auto-scaling Business Model

- User: Specify workload and min/max memory requirements
- Service provider: Run it somewhere, charge as they go
- Achieving high performance and low price is the provider’s duty



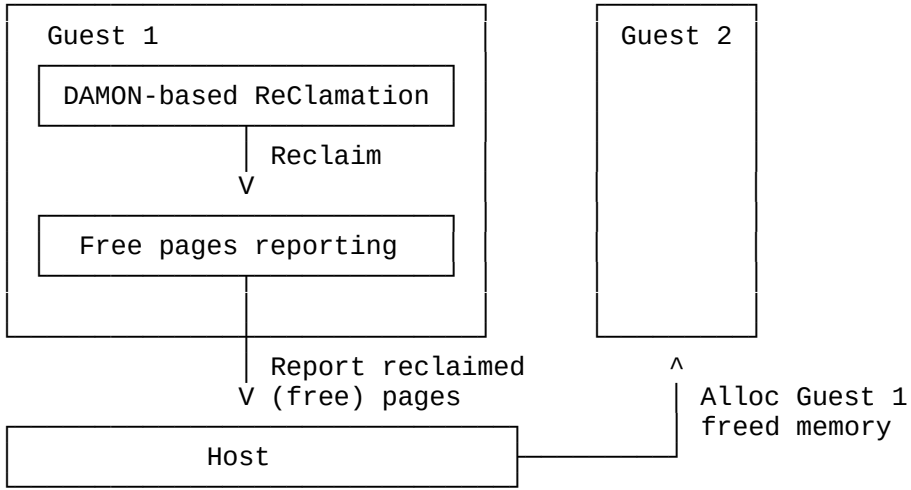
Designs of Collaborative Memory Auto-scaling

- Non-collab approach: Host reclaims/reallocs guests' memory
- Collab approach: Host reallocs guests-reported free pages
- Minimize perf degradation from host's mistakes
- Question: What if guests are not memory frugal?



Design of AWS' Collaborative Memory Auto-scaling

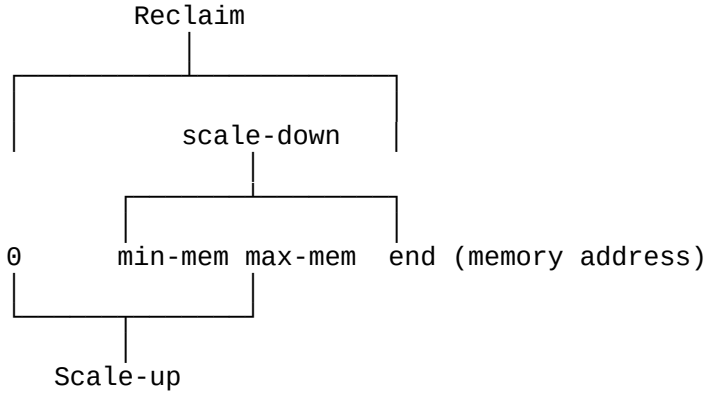
- Guests report free pages to host; the host re-alloc reported pages
- Guests run DAMOS-based proactive reclamation for increasing free pages



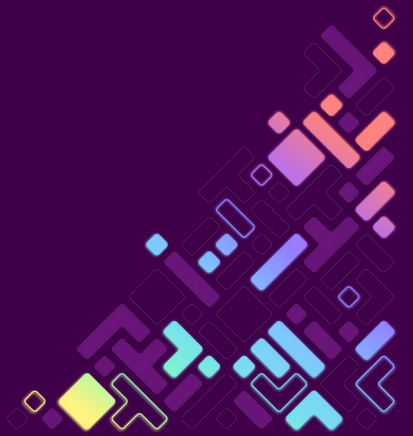
AWS' DAMON-based Proactive Reclamation

- AWS Aurora Serverless v2 is an officially known user of the memory scaling
- Holistic re:Invent of the entire memory auto-scaling with DAMON is ongoing

[RFC IDEA v2 0/6] mm/damon: introduce Access/Contiguity-aware Memory Auto-scaling (ACMA)

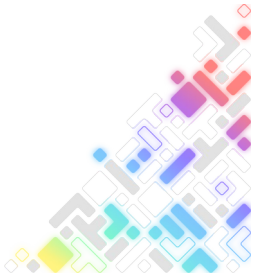


CXL Memory Tiering

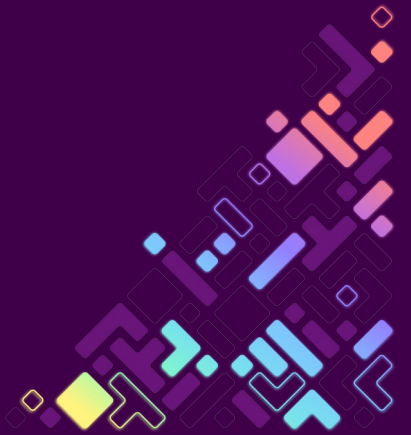


Moving Mic to Honggyu Kim

- Hongguu Kim from SK hynix will introduce this recipe
- If you are reading these slides after the talk, refer to Honggyu’s slides (damon_recipes_osseu_hmsdk.pdf)

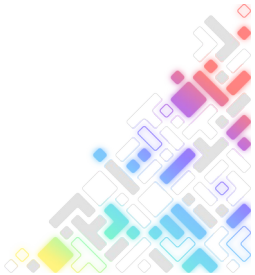


DAMON Community



DAMON Owner: Community

- DAMON is a community-driven development project
- Not owned by individual or companies (e.g., SJ, Amazon, Meta)
- Owned by the community
- A number of DAMON features are results of community discussions



Mailing List

- Primary and official place for the community
- For any question, discussion, reports, and patches
- A tool (hkml) for mailing list-unfamiliar people is supported: https://github.com/sjp38/hackermail
- Mailing address: damon@lists.linuxdev
- Archive: https://lore.kernel.org/damon
- Don’t hesitate sending personal mails/DMs to the maintainer (SJ)

Community Meetups: Beer/Coffee/Tea Chats

- For *any* informal chats
- Regular bi-weekly open/registration-based discussions
- Occasional/regular private meetings on demand
- Feel free to join existing meeting series, or ask one for yours
- Multiple individuals and companies are using these
- Google doc and calendar for schedules are available

Generated from
https://qr.io

Project Website

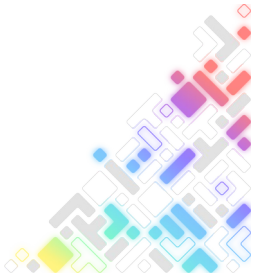
- https://damonitor.github.io
- Provides all information for DAMON starters including news, demo, usage guides

Table Of Contents

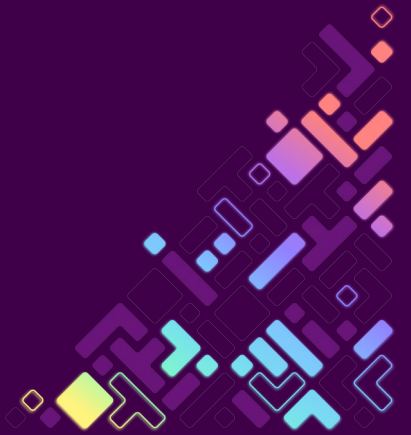
Generated from https://qr.io

Call For Your Voices

- We prefer random evolution over intellectual design
- Put your voice on the evolution path for *your* purposes
- Report your use case, test results, and challenges you gone through
- Ask questions and request features/documents
- Show your interest to known TODO items
- Send patches
- The maintainer is committed to help the community, you



Summary



Summary

- DAMON is an efficient data access pattern snapshot generator
- People can optimize system memory usage using DAMON
- AWS Aurora Serverless v2 and SK hynix HMSDK are using DAMON for memory auto-scalaing and CXL memory tiering, respectively
- DAMON community is waiting for your voices
- If you want to start, visit the project site: https://damonitor.github.io

Generated from
https://qr.io

Questions?

- If your question is not answered by the session, use below
- The maintainer: sj@kernel.org
- Project webpage: https://damonitor.github.io
- Kernel docs for admin and kernel programmers
- DAMON mailing list: damon@lists.linux.dev
- DAMON Beer/Coffee/Tea Chat
- Today’s in-person DAMON meetup (4:55 pm, room 444)

Generated from https://qr.io

THE LINUX FOUNDATION

OPEN SOURCE SUMMIT

EUROPE