Mmap Advise

Table of contents

Alternative Datapath Options

Samples

Cache Invalidate Sample

Introduction

DOCA Mmap Advise is used to give advanced memory-related instructions to NVIDIA® BlueField® DPUs in order to improve system or application performance.

> **Note**
>
> To use DOCA Mmap Advise with BlueField, the device must be configured to work in DPU mode as described in BlueField Modes of Operation.

The operations in the instructions are meant to influence the performance of the application, but not its semantics. The operations allow an application to inform the NIC how it expects it to use some mapped memory areas, so the BlueField's hardware can choose appropriate optimization techniques.

Prerequisites

DOCA Mmap Advise is a context and follows the architecture of a DOCA Core Context, it is recommended to read the following sections of the DOCA Core page before proceeding:

- DOCA Core Execution Model
- DOCA Core Device
- DOCA Core Memory Subsystem

Architecture

DOCA Mmap Advise is a DOCA Context as defined by DOCA Core. S ee DOCA Core Context for more information.

DOCA Mmap Advise currently supports the following list of advised operations:

- Cache Invalidate Operation

Cache Invalidate Operation

When data is processed by BlueField's cores it may be temporarily stored in the cores' system-level cache (i.e., L3 cache). When a cache line is occupied and new data must be written to it, the cache management sub-system evicts the existing data, usually based on LRU policy, by performing a write-back operation to store this data in the main (DDR) memory. When this data is not required to be stored in the BlueField's memory (e.g., it is host data and is no longer needed after it is copied to the host's memory), the cache's write-back operation wastes memory bandwidth that reduces overall system performance, which is undesirable. The simplest to avoid this write-back operation is to mark the appropriate cache lines as "invalid". This enables their immediate reuse, without additional operations.

The cache invalidate operation facilitates invalidating a set of cache lines.

Environment

Applications based on DOCA Mmap Advise can run on the BlueField target.

Objects

Device and Device Representor

The MMAP Advise context requires a DOCA Device to operate. The device is used to access memory and perform the copy operation. See DOCA Core Device Discovery.

Info

For the same DPU, it does not matter which device is used (i.e., PF, VF, SF) as all these devices utilize the same hardware components.

Note

The device must stay valid for as long as the MMAP Advise instance is not destroyed.

Memory Buffers

The cache invalidate task requires one DOCA Buffer containing the address space to invalidate depending on the allocation pattern of the buffers (refer to the table in section "Inventory Types"). To find what kind of memory is supported, refer to the table in section "Buffer Support".

Buffers must not be modified or read during the cache invalidate operation.

Configuration Phase

To start using the context, users must go through a configuration phase as described in DOCA Core Context Configuration Phase.

This section describes how to configure and start the context, to allow execution of tasks and retrieval of events.

Configurations

The context can be configured to match the application's use case.

To find if a configuration is supported, or what the min/max value for it is, refer to section "Device Support".

Mandatory Configurations

These configurations are mandatory and must be set by the application before attempting to start the context:

- At least one task/event type must be configured. See configuration of tasks and/or events in sections "Tasks" and "Events" respectively for information.
- A device with appropriate support must be provided upon creation

Device Support

DOCA Mmap Advise requires a device to operate. To pick a device, refer to DOCA Core Device Discovery.

As device capabilities may change (see DOCA Core Device Support), it is recommended to select your device using the following method:

- `doca_mmap_advise_cap_task_cache_invalidate_is_supported`

Some devices expose different capabilities as follows:

- Maximum cache invalidate buffer size may differ.

Buffer Support

Tasks support buffers with the following features:

Buffer Type	Buffer
Local mmap buffer	Yes
MMAP from PCIe export buffer	No
MMAP from RDMA export buffer	No
Linked list buffer	No

Execution Phase

This section describes execution on the CPU using DOCA Core Progress Engine.

Tasks

DOCA Mmap Advise exposes asynchronous tasks that leverage DPU hardware according to the DOCA Core architecture. See DOCA Core Task for information.

Cache Invalidate Task

The cache invalidate task facilitates invalidating a set of cache lines, preventing them from being written back to the RAM (thus increasing performance).

Task Configuration

Description	API to Set the Configuration	API to Query Support
Enable the task	`doca_mmap_advise_task_invalidate_cache_set_conf`	`doca_mmap_advise_cap_task_cache_invalidate_is_supported`
Number of tasks	`doca_mmap_advise_task_invalidate_cache_set_conf`	–
Maximal buffer size	–	`doca_mmap_advise_task_cache_invalidate_get_max_buf_size`
Maximal buffer list size	–	–

Task Input

Common input as described in DOCA Core Task.

Name	Description
buffer	Buffer that points to the memory to be invalidated

Task Output

Common output as described in DOCA Core Task.

Task Completion Success

After the task is completed successfully:

- The cache is invalidated

Task Completion Failure

If the task fails midway:

- The context may enter stopping state, if a fatal error occurs
- The cache is not invalidated

Task Limitations

- The operation is not atomic
- Once the task has been submitted, the buffer should not be read/written to
- Other limitations are described in DOCA Core Task

Events

DOCA Mmap Advise exposes asynchronous events to notify on changes that happen unexpectedly, according to DOCA Core architecture.

The only events DOCA Mmap Advise exposes are common events as described in DOCA Core Event.

State Machine

DOCA Mmap Advise context follows the context state machine as described in DOCA Core Context State Machine.

The following section describes how to move states and what is allowed in each state.

Idle

In this state it is expected that the application:

- Destroys the context
- Starts the context

Allowed operations:

- Configuring the context according to section "Configurations"
- Starting the context

It is possible to reach this state as follows:

Previous State	Transition Action
None	Create the context
Running	Call stop after making sure all tasks have been freed
Stopping	Call progress until all tasks are completed and freed

Starting

This state cannot be reached.

Running

In this state, it is expected that the application:

- Allocates and submits tasks
- Calls progress to complete tasks and/or receive events

Allowed operations:

- Allocating a previously configured task
- Submitting a task
- Calling stop

It is possible to reach this state as follows:

Previous State	Transition Action
Idle	Call start after configuration

Stopping

In this state it is expected that the application:

- Calls progress to complete all in-flight tasks (tasks complete with failure)
- Frees any completed tasks

Allowed operations:

- Call progress

It is possible to reach this state as follows:

Previous State	Transition Action
Running	Call progress and fatal error occurs
Running	Call stop without freeing all tasks

Alternative Datapath Options

DOCA Mmap Advise only supports datapath on the CPU. See section "Execution Phase".

Samples

Cache Invalidate Sample

The sample illustrates how to invalidate the cache for a memory range after copying it using DOCA DMA.

The sample logic includes:

1. Locating DOCA device.
2. Initializing needed DOCA core structures.
3. Populating DOCA memory map with two relevant buffers.
4. Allocating element in DOCA buffer inventory for each buffer.
5. Initializing DOCA DMA memory copy task object.
6. Initializing DOCA Mmap Advise cache invalidate task object
7. Submitting DMA task.
8. Polling for completion:
    1. Handling DMA task completion – submitting the cache invalidate task in the DMA task completion callback body.
    2. Handling cache invalidate task completion.
9. Polling for completion.
10. Destroying DMA, DOCA MMAP Advise, and DOCA Core objects.

Reference:

- `/opt/mellanox/doca/samples/doca_common/cache_invalidate/cache_inv`
- `/opt/mellanox/doca/samples/doca_common/cache_invalidate/cache_inv`
- `/opt/mellanox/doca/samples/doca_common/cache_invalidate/meson.bui`