NVM Express® TP4146a Flexible Data Placement

Overview

Presented by
Mike Allison (Samsung) and John Rudelic (Solidigm)
FDP Architecture
An FDP configuration consists of:
- One or more Reclaim Units (RUs)
- One or more Reclaim Groups (RGs)
- One or more Reclaim Unit Handles (RUHs) that reference to a Reclaim Unit in each RG

An Endurance Group may supports one or more FDP configurations

Write commands specify which RU to write the logical blocks by specifying:
- An RUH
- An RG
Namespace Creation

- Namespace creation using the Namespace Management command
  - Specify the Endurance Group with FDP enabled
  - Host may provide a Placement Handle List

<table>
<thead>
<tr>
<th>Placement Handle</th>
<th>Reclaim Unit Handle Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>$N-1$</td>
</tr>
</tbody>
</table>
## Log Pages

<table>
<thead>
<tr>
<th>FDP Log Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDP Configurations</td>
</tr>
<tr>
<td>Reclaim Unit Handle Usage</td>
</tr>
<tr>
<td>FDP Statistics</td>
</tr>
<tr>
<td>FDP Events</td>
</tr>
</tbody>
</table>
### Log Pages – FDP Configurations

<table>
<thead>
<tr>
<th>FDP Configuration Log Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>#FDP Configurations (N)</td>
</tr>
<tr>
<td>Version</td>
</tr>
<tr>
<td>Size</td>
</tr>
<tr>
<td>FDP Config Descriptor 0 (FDPCD0)</td>
</tr>
<tr>
<td>FDP Config Descriptor 1 (FDPCD1)</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>FDP Config Descriptor N (FDPCDN)</td>
</tr>
</tbody>
</table>

#### FDP Configuration Descriptor
- Descriptor Size
- FDP Attributes
  - Vendor Specific Size (VSS)
  - # RGs (NRG)
  - # RU Handles (NRUH)
  - MAX Placement IDSs (MAXPIDS)
  - # Namespaces supported
  - Reclaim Unit Nominal Size (RUNS)
  - Estimated RU Time Limit (ERUTL)
  - RUH Descriptor List (RUHD0)
  - RUH Descriptor List (RUHD1)
  - RUH Descriptor List (RUHDN)

#### FDP Attributes
- FDP Config Valid
- FDP Volatile Write Cache (FDPVWC)
- RG ID Format (RGIF)

#### RUH Descriptor
- Handle type (initial or persistent)
- Vendor specific
# Log Pages – FDP Reclaim Unit Handle Usage

<table>
<thead>
<tr>
<th>FDP Reclaim Unit Handle Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Reclaim Unit Handles (NRUH)</td>
</tr>
<tr>
<td>RUH Usage Descriptor 0 (RUHUD0)</td>
</tr>
<tr>
<td>RUH Usage Descriptor 1 (RUHUD1)</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>RUH Usage Descriptor N (RUHUDN)</td>
</tr>
</tbody>
</table>

**RUH Attributes**
Controller/Host specified
Log Pages – FDP Statistics

<table>
<thead>
<tr>
<th>FDP Statistics Log Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Bytes with Metadata Written (HBMW)</td>
</tr>
<tr>
<td>Media Bytes with Metadata Written (MBMW)</td>
</tr>
<tr>
<td>Media Bytes Erased (MBE)</td>
</tr>
</tbody>
</table>
Log Pages – FDP Events

FDP Events Log Page

#FDP Events (N)
FDP Event 1
FDP Event 2
...
FDP Event N

FDP Event

FDP Event Type
FDP Event Flags (FDPEF)
Placement Identifier (PID)
Event Timestamp
Namespace Identifier (NSID)
Event Type Specific
Reclaim Group Identifier
Reclaim Unit Handle Identifier
Vendor Specific Info

FDP Event Type
HOST EVENTS
RU not fully written to capacity
RU active3 time limit exceeded
Controller level reset – modified RUH
Invalid PID
Vendor specific
CONTROLLEER EVENTS
Media reallocated
Implicitly modified RUH
Vendor Specific

FDP Event Flags (FDPEF)
Location Valid (LV)
NSID Valid (NSIDV)
Placement ID (PIV)
Configuring & Monitoring
#Check NVMe Identify Controller Data Structure
if (FDP Support Bit = 1)  #FDP Supported
    #Configure FDP
    #Read FDP Log Page for FDP configurations
    #Set feature – enable FDP configuration
    #(Optional) Configure Placement handle list
    #(Optional) Enable FDP events
    #Set feature – enable FDP set FDPE=1
else #FDP Not supported
# FDP Monitoring Loop
# Check FDP Events
    # respond to event
# Check FDP RUH
    # respond to RUH
# Check FDP Statistics
    # respond to Statistics
FDP Writing Command Examples
Host issues a Write command to Namespace A specifying Placement Handle 1 and Reclaim Group 0.
Controller looks up Placement Handle 1 in Namespace A and determines to place the data using Reclaim Unit Handle 2 on Reclaim Group 0.
Controller looks up Placement Handle 1 in Namespace A and determines to place the data using Reclaim Unit Handle 2 on Reclaim Group 0.
Controller uses Reclaim Unit Handle 2 to place the data in Reclaim Unit 3 of Reclaim Group 0.
Host issues a Write command to Namespace A specifying Placement Handle 2 and Reclaim Group 0
Controller looks up Placement Handle 2 in Namespace A and determines to place the data using Reclaim Unit Handle 3 on Reclaim Group 0.
Write to fill the capacity of a Reclaim Unit

Controller uses Reclaim Unit Handle 3 to place the data in Reclaim Unit 4 in Reclaim Group 0
Write to fill the capacity of a Reclaim Unit

Controller detects Reclaim Unit 4 is written to capacity and modifies Reclaim Unit Handle 3 to reference the empty Reclaim Unit 5

Controller

Namespace A

Placement Handle | Reclaim Unit Handle Identifier
---|---
0 | 0
1 | 2
2 | 3

SSD – Endurance Group 1

Reclaim Group 0
- Reclaim Unit 0
- Reclaim Unit 1
- Reclaim Unit 2
- Reclaim Unit 3
- Reclaim Unit 4
- Reclaim Unit 5
Host issues a Write command to Namespace A specifying Placement Handle 1 and Reclaim Group 0
Controller looks up Placement Handle 1 in Namespace A and determines to place the data using Reclaim Unit Handle 2 on Reclaim Group 0.
Write to fill the capacity of a Reclaim Unit

Controller uses Reclaim Unit Handle 2 to place a portion of the data in Reclaim Unit 3 in Reclaim Group 0 filling Reclaim Unit 3 to capacity.
Write to multiple Reclaim Units

Controller Reclaim Unit Handle 2 to reference the empty Reclaim Unit 2

SSD – Endurance Group 1
Controller uses Reclaim Unit Handle 2 to place the remaining portion of the data in Reclaim Unit 2 in Reclaim Group 0.

SSD – Endurance Group 1
Backwards Compatible Writes

Host issues a Write command to Namespace A without specifying a Placement Handle or an Reclaim Group

SSD – Endurance Group 1

Controller

Namespace A

Placement Handle | Reclaim Unit Handle Identifier
---|---
0 | 0
1 | 2
2 | 3

Reclaim Unit 0
Reclaim Unit 1
Reclaim Unit 2
Reclaim Unit 3
Reclaim Unit 4
Reclaim Unit 5
Controller looks up Placement Handle 0 in Namespace A and determines Reclaim Unit Handle 0 is used for the write and selects Reclaim Group 0.
Controller uses Reclaim Unit Handle 0 to place the data in Reclaim Unit 0 in Reclaim Group 0.
Comparing NVMe® Data Placements

Streams
Flexible Data Placement (FDP)
Zoned Namespaces (ZNS)
<table>
<thead>
<tr>
<th>Streams</th>
<th>FDP</th>
<th>ZNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-conforming writes are not logged</td>
<td>Non-conforming writes are logged</td>
<td>Error on Write</td>
</tr>
<tr>
<td>Known alignment only after format</td>
<td>Commands available to host to stay aligned</td>
<td>Always aligned by interface rules</td>
</tr>
<tr>
<td>WAF = 1 achievable without feedback</td>
<td>WAF = 1 achievable with feedback</td>
<td>WAF = 1 guaranteed</td>
</tr>
<tr>
<td>Backwards compatible</td>
<td>Backwards compatible</td>
<td>Not Backwards compatible</td>
</tr>
<tr>
<td>No information that controller moved user data</td>
<td>Post logging that controller moved user data</td>
<td>Notification for Host to move user data</td>
</tr>
<tr>
<td>Placement identifier not tied to LBA</td>
<td>Placement identifier not tied to LBA</td>
<td>Placement identifier is the LBA</td>
</tr>
<tr>
<td>Stream Granularity Size (SGS)</td>
<td>Reclaim Units</td>
<td>Zones</td>
</tr>
<tr>
<td>SGS capacity = SGS Size</td>
<td>Reclaim Unit capacity = Reclaim Unit size</td>
<td>Zone capacity &lt;= Zone size</td>
</tr>
<tr>
<td>No Host metadata per SGS</td>
<td>No Host metadata per Reclaim Unit</td>
<td>Host metadata per Zone</td>
</tr>
<tr>
<td>Namespace capacity defines # SGS</td>
<td>Endurance Group capacity defines # Reclaim Units</td>
<td>Namespace Capacity defines # zones</td>
</tr>
<tr>
<td>Sequential, Random, and Over Write</td>
<td>Sequential, Random, and Over Write</td>
<td>Sequential Write</td>
</tr>
<tr>
<td>Writes allowed to cross Boundaries</td>
<td>Writes allowed to cross Boundaries</td>
<td>Writes not allowed across Boundaries</td>
</tr>
<tr>
<td>QD &gt; 1: LBA known at Write Submission</td>
<td>QD &gt; 1: LBA known at Write Submission</td>
<td>QD &gt; 1: LBA known at Write Completion Zone Append command</td>
</tr>
<tr>
<td>Stream written by a single namespace</td>
<td>Reclaim Unit written by one or more namespaces</td>
<td>Zone written by a single namespace</td>
</tr>
<tr>
<td>API is Stateless</td>
<td>API is Stateless</td>
<td>API is Stateful</td>
</tr>
<tr>
<td>Requires Full FTL Table</td>
<td>Requires Full FTL Table</td>
<td>Full FTL Table not required</td>
</tr>
<tr>
<td>Dynamic write resource allocation</td>
<td>Static write resource allocation</td>
<td>Dynamic write resource allocation</td>
</tr>
</tbody>
</table>
Please take a moment to rate this session.

Your feedback is important to us.