Storage Data Placement TWG Proposal

October-27-2023

1 Charter Submission

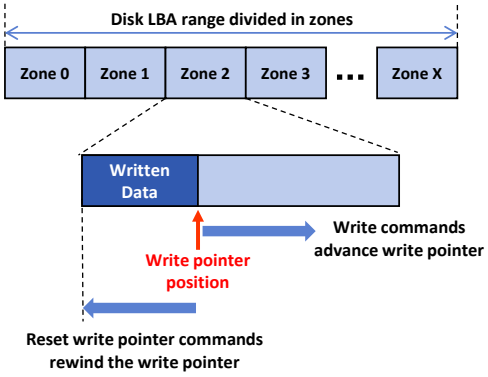
1.1 Preamble

TITLE: Storage Data Placement TWG
SUBMITTER: Zoned Storage TWG
DATE SUBMITTED: 27 October 2023
VERSION: 1.1

1.2 Charter

Storage data placement is becoming a fundamental building block of today's storage ecosystem. Out of several storage data placement methods, two examples that are being developed are: Zoned Storage and Flexible Data Placement. In both of these storage ecosystems, the host contributes to storage data placement on the device. For SSDs, this host/device cooperation enables lower overprovisioning (i.e., delivering more media to the end user), and more predictable performance. Device costs may also be reduced by enabling less device-based DRAM. For HDDs, Zoned Storage enables access to SMR drives, which deliver the highest capacities.

In a Zoned Storage system, the logical block address space of a storage device is divided into defined regions, called zones, and various rules are applied to I/O operations in those regions. The most common rule applied in a zoned storage system is that the host must write data sequentially to a zone or the device will reject the command. There are other zoned storage use cases in which the zone access rules allow varying amounts of random I/O and/or overwrites within zones.



In a Flexible Data Placement (FDP) system, the host specifies grouping of data that is expected to be managed at the same time. The interfaces for this and the parameters that control this are defined in the NVM Express® NVM Command Set specification. There is a need to define how those parameters should be used for optimal performance of the storage device and a general description of how applications can use FDP to achieve the best Write Amplification Factor (WAF) from a device.

While command interfaces for Zoned Storage and Flexible Data Placement have been standardized (ZAC/ZBC for SMR HDDs and ZNS/FDP for NVMe SSDs), the command specifications themselves leave substantial flexibility regarding how host software interacts with the storage I/O stack and storage devices. For example:

- there are use cases in which only a small number of very large zones are preferred, and use cases where many small zones are preferred.
- there are use cases where a device is formatted entirely for zoned storage, and use cases where a device may contain both zoned storage and conventional block storage.

Due to the diversity of use cases for storage data placement, and the corresponding diversity of host/device interaction models, the data placement storage ecosystem would benefit from the descriptions of common use cases, and a nomenclature around which corresponding host/device models can be described. Such a descriptions will provide guidance to the SW community (from host

application to middleware to kernel/file systems) and to device vendors on how to design storage data placement building blocks. This, in turn, will provide a foundation for the development of a broad based and interoperable storage data placement SW and HW ecosystem.

Proposed Charter of Storage Data Placement TWG
Facilitate a common industry understanding of storage data placement use cases and host/device architecture and programming models, providing a framework for storage data placement SW and HW design, and enabling the development of a robust storage data placement ecosystem.

2 Program of Work

2.1 Preamble

TITLE: Storage Data Placement TWG
SUBMITTER: Zoned Storage TWG
DATE SUBMITTED: 27 October 2023
VERSION: 1.1

2.2 Proposed Program of Work

The proposed Storage Data Placement TWG will focus on the following main areas of work:

1) Zoned Storage use case white papers;
2) Zoned Storage Architecture and Programming Model specification; and
3) Flexible Data Placement use case white papers.

Use Cases/IO Patterns/Workloads

The ecosystem needs a clearer understanding of the use cases and workloads which are most applicable to storage data placement solutions. The TWG will document key use case examples and their characteristics in white paper form. Some use case examples are as follows, though this list will likely change/evolve as the TWG does its work:

1. Write-Once, Read-Many Workloads
    - This case covers storage for primarily immutable data, such as Archival, Video Surveillance, and Video Streaming.
2. IO Intensive
    - This use case covers use of Zoned Storage devices to serve high I/O workloads, such as database systems and filesystems. Specifically, log-structured and copy-on-write workloads may be easily adapted to zoned storage. In this use case, zones are generally mapped to existing objects, reducing WAF.
3. Tiered Storage
    - This case covers use of Zoned Storage devices in tiered systems, such as all-flash arrays and distributed filesystems, in which data is cached in a faster storage tier, where the data is written randomly before being later written sequentially to a specific zoned storage device tier.

Host/Device Model and Programming Model

The TWG will develop specifications, informed by use cases, which describes the general host/device programming model for storage data placement. These specifications will describe, among other things, the general mechanisms for how Zoned Storage and Flexible Data Placement are manifested in a system. For example:

1) The Zoned Storage Host/Device Model will include items such as that the SSD has the sole responsibility for managing reliability of the media, including dynamic wear-leveling of media associated to zones.
2) The Zoned Storage Programming Model will include items such as how a zoned storage device can be accessed through common storage stacks, such as the general Linux storage stack and SPDK.
3) The FDP white paper will describe:
    a. how characteristics that a device reports are able to be utilized by various types of applications;
    b. how characteristics that can be configured on that device are able to be utilized by various types of applications; and
    c. best practices for utilizing FDP devices.

This program of work is not intended to define new storage data placement device standards, as this work takes place within NVMe, T10, and T13, etc. Neither does the TWG intend to develop software. It is expected, however, that the work of this TWG will facilitate the development of storage data placement solutions ecosystem.