Flash Memory Summit Session:

Benefits of ZNS in Datacenter Storage Systems

Woosuk Chung, Director, Memory Systems R&D
Legal Disclaimer

The information contained in this document is claimed as property of SK hynix. It is provided with the understanding that SK hynix assumes no liability, and the contents are provided under strict confidentiality.

This document is for general guidance on matters of interest only. Accordingly, the information herein should not be used as a substitute for consultation or any other professional advice and services.

SK hynix may have copyrights and intellectual property right. The furnishing of document and information disclosure should be strictly prohibited.

SK hynix has right to make changes to dates, product descriptions, figures, and plans referenced in this document at any time. Therefore the information herein is subject to change without notice.

© 2019 SK hynix Inc. All rights reserved
CONTENTS

Introduction
Performance Evaluation
RocksDB, F2FS with ZNS SSD
Summary
Introduction: Zoned Namespaces Proposal

- Previous proposals are not a complete solution for data center storage system
- New proposal, Zoned Namespaces (ZNS), appears to be an optimal solution

<table>
<thead>
<tr>
<th></th>
<th>Log Abstraction</th>
<th>In-Host Placement Policy</th>
<th>In-Drive Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Streams SSD (HotStor `14)</td>
<td>×</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>OCSSD 1.2 (ASPLOS `16)</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>IO Determinism (Fall `16)</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>OCSSD 2.0 (FAST `17)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zoned Namespaces (NVMe Spec. `19)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Reference: MSFT, `17 Storage Developer Conference
Introduction: Case of Multi-Streams

- Garbage Collection is reduced but not completely removed
- Lifespan & Performance can be enhanced but not to the optimal level
Introduction: ZNS Concept

- Only sequential write is accepted in each Zone (random write is not allowed)
- Zones are erased by the host issuing a special command, Zone Reset (no GC)
Introduction: SK hynix’s ZNS SSD Prototype

- Prototype available on two SK hynix SSD products

<table>
<thead>
<tr>
<th>Item</th>
<th>PE4011</th>
<th>PE6011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface</td>
<td>PCIe Gen3 x 4</td>
<td>PCIe Gen3 x 4</td>
</tr>
<tr>
<td>Protocol</td>
<td>NVMe 1.2.1</td>
<td>NVMe 1.3</td>
</tr>
<tr>
<td>Form Factor</td>
<td>M.2 22110</td>
<td>U.2 7mm</td>
</tr>
<tr>
<td>Capacity</td>
<td>1920GB</td>
<td>3840TB</td>
</tr>
<tr>
<td>NAND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>512 Gb</td>
<td>512 Gb</td>
</tr>
<tr>
<td>Type</td>
<td>3D TLC</td>
<td>3D TLC</td>
</tr>
</tbody>
</table>
Performance Evaluation: Configuration & Environment

- New ZNS commands added in SPDK
- Emulated workload generated and run by FIO
- Kernel S/W stack is bypassed to remove overhead

**Test Framework**

User

<table>
<thead>
<tr>
<th>FIO</th>
<th>Zoned Block Device</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SPDK

<table>
<thead>
<tr>
<th>ZNS I/O Engine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Kernel

<table>
<thead>
<tr>
<th>Bypass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Storage

<table>
<thead>
<tr>
<th>ZNS / Conventional SSD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Environment**

- **Hardware:**
  - CPU: Intel(R) Xeon(R) CPU E5-2667 v4 @ 3.20GHz
  - Memory: 251GB
  - SSD(ZNS/Conventional): PE4011, PE6011

- **Software:**
  - Ubuntu 15.04
  - Linux 4.20.0 x86_64
  - FIO-3.12, SPDK
Performance Evaluation: Expected Benefits

I. Extend SSD Lifespan
II. Reduce Read Tail Latency (QoS)
III. Improve I/O Performance
IV. Reduce Overprovisioning
V. Reduce DRAM in SSD
Performance Evaluation: Expected Benefits (1/5)

- Extend SSD Lifespan
  - Increase lifespan 3x for the case of 8-writes
  - No Garbage Collection is required

\[ WAF = \frac{\text{Bytes written to NAND}}{\text{Bytes written from Host}} \]
Performance Evaluation: Expected Benefits (2/5)

- Improve Read Tail Latency
  - Reduce IO interference by SSD’s internal background operations
  - Improve read response for mixed workloads
Performance Evaluation: Expected Benefits (3/5)

- Improve I/O Performance
  - Getting consistent throughput
  - Higher bandwidth for mixed workloads (1 Random Read and 5 Writes)
Performance Evaluation: Expected Benefits (4/5)

- Less Overprovisioning (OP)
  - Reduce OP area for removing Garbage Collection
  - Eventually, user capacity is increased
Performance Evaluation: Expected Benefits (5/5)

- Reduce DRAM size requirement for SSD
  - Less DRAM is required per 1TB capacity
  - Make rooms for more critical DRAM use
SW Enablement: RocksDB, F2FS with ZNS SSD

- “Linux kernel that supports Zoned Device” + “Two types of PE6011 SSD”
  - PE6011-based ZNS SSD for append-only write in F2FS
  - Conventional SSD for random-write in F2FS
SW Enablement: Live Demo

- Running RocksDB with db_bench on ZNS-configured Linux Host System
- Key-value data is stored into PE6011 ZNS SSD
Summary

1. Two SK hynix’s NVMe ZNS Prototypes
   - PE4011
   - PE6011

2. Expected Benefits with ZNS
   - Extend SSD Lifespan
   - Improve Performance & QoS
   - Reduce SSD’s memory resource requirement (OP, DRAM)

3. Linux Host enabled with ZNS Prototype
Learn more about SK hynix

Visit SK hynix @ booth #407

Experience SK hynix products and demos & get a free giveaway!
Thank you

Growing together for better tomorrow