Position: Synergetic Effects of Software and Hardware Parameters on the LSM System

Authors:

Jinghuan Yu, Heejin Yoon*

Sam H. Noh*, Young-ri Choi*, Chun Jason Xue

香港城市大學
City University of Hong Kong

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

*

Log Structured Merge-tree (LSM)

Specific designs for HDD and write-intensive workload.

Does the working principle of LSM still fit these new mediums?

Periodical compaction with various resource occupation.

What is the critical factor deciding performance?

Performance Feature of Each Media

SATA SSD

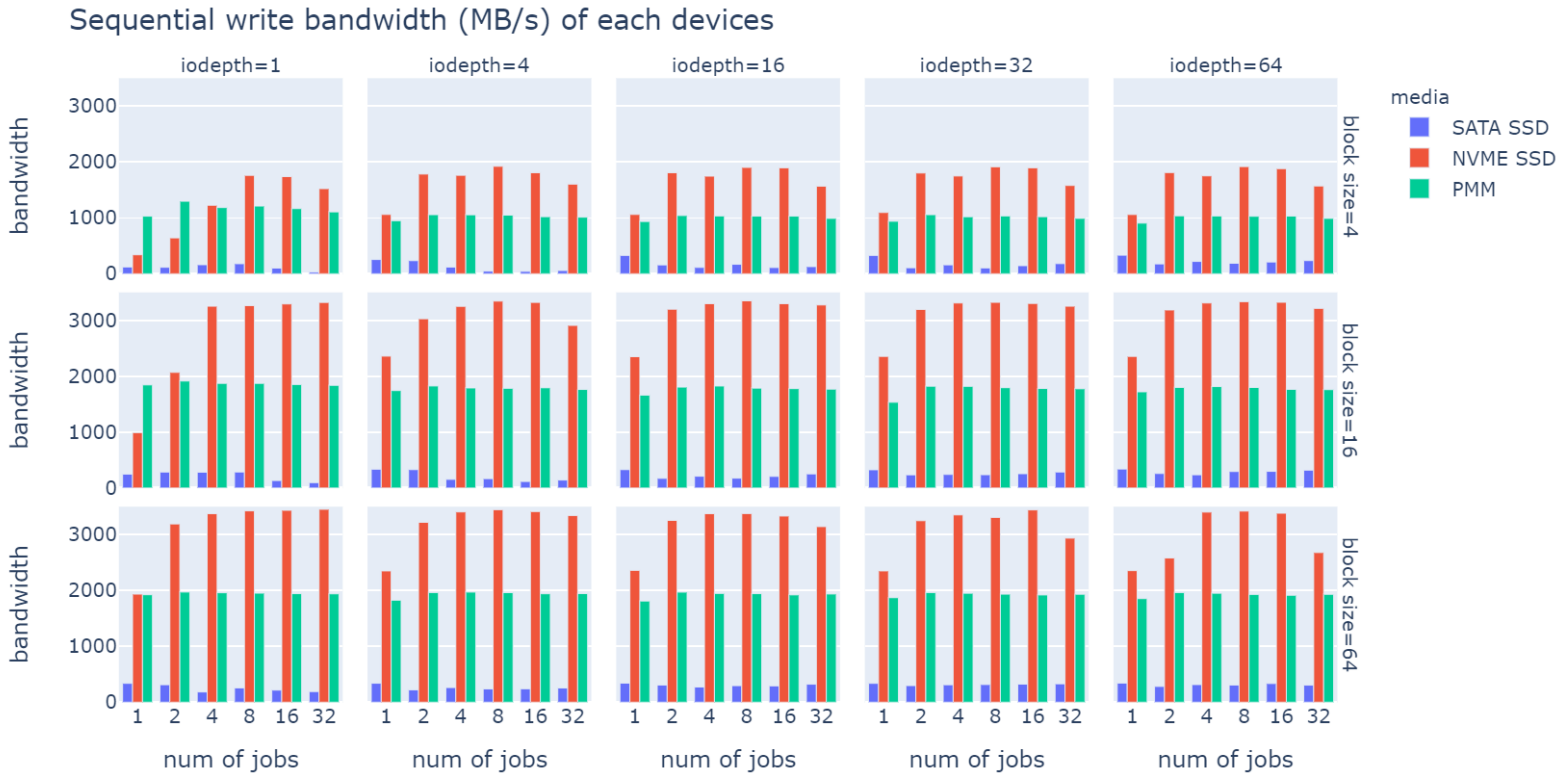
- Limited bandwidth
- Poor parallel performance
- Unstable performance due to foreground garbage collection

NVMe SSD

- Highest sequential bandwidth
- Strong parallel performance, with higher requirement for CPU
- Performance is strongly affected by write granularity

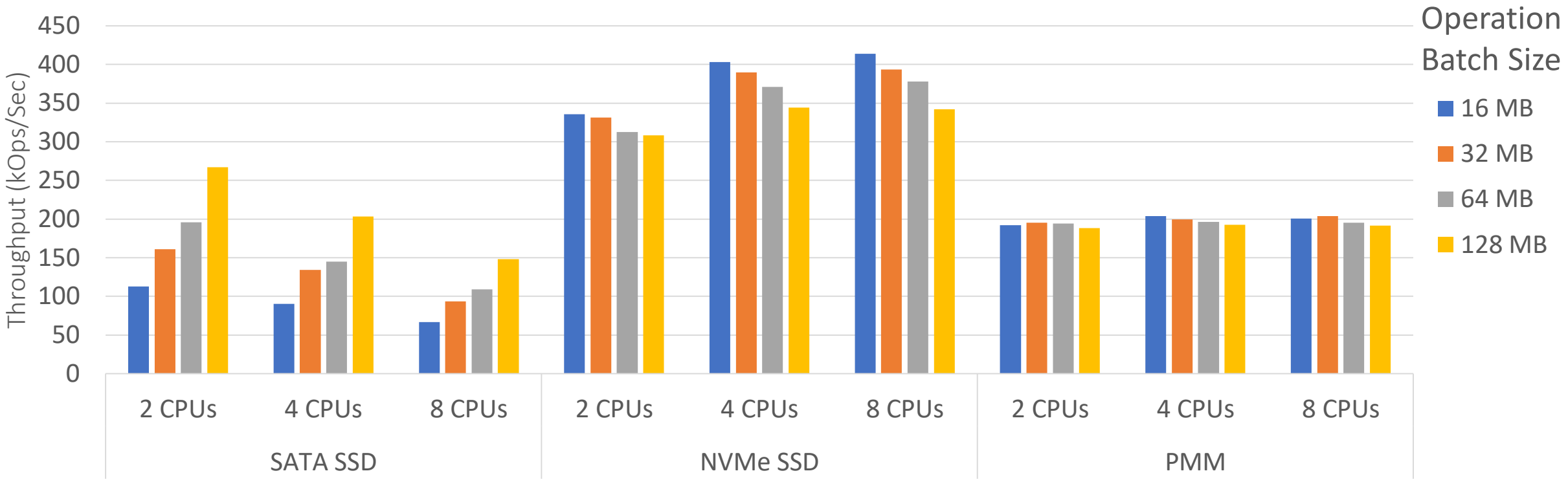
PMM

- Relatively low latency
- Stable performance in any parallelism
- Strong wear lifetime without GC



Media Type	Average Access Latency (µs)
SATA SSD	37.78
NVMe SSD	11.77
PMM	2.61

Performance Comparison of Devices in RocksDB



- Increasing number of CPUs causes IO congestion, decreasing performance
- With fixed CPUs, benefits from larger batch size

- Best throughput
- Performance increase tends to be stable as the number of CPUs increases
- Suffers from larger batch size

- Not sensitive to the number of CPUs or batch size
- Throughput difference is far from bandwidth comparison

Existing Solutions		Our Targets
• Rule-based selection • Size-based scaling	Heterogenous Storage	• Auto scaling • Driven by workload
• Unified configuration • Offline tuning • Based on statistics data	Parameter Tuning	• Device oriented • Online tuning • Based on quantitative analysis
• Disk utilization first • Lazy scheduling • Devices features based	Resource Utilization	• Both CPU and disk utilization • Smooth, effective, and predictable

VS

Characteristic analysis and design points

Performance Traits of SATA SSD

Strength

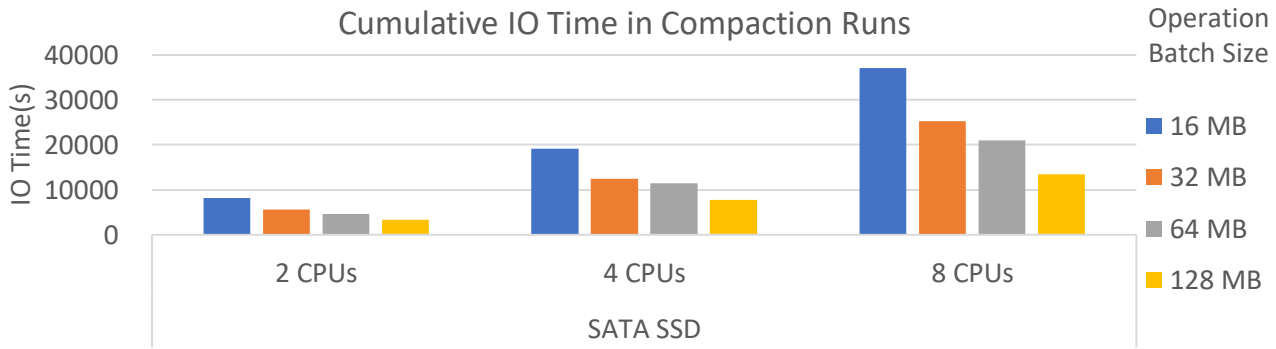
- Effective for bulk single-thread write workload

Weakness

- Serious IO congestion during multi-thread compaction

Design Opportunities

- Single queue continuous write
- Large-grained operations



In SATA SSD, IO time increases dramatically as number of CPUs increases

Performance Traits of NVMe SSD

Strength

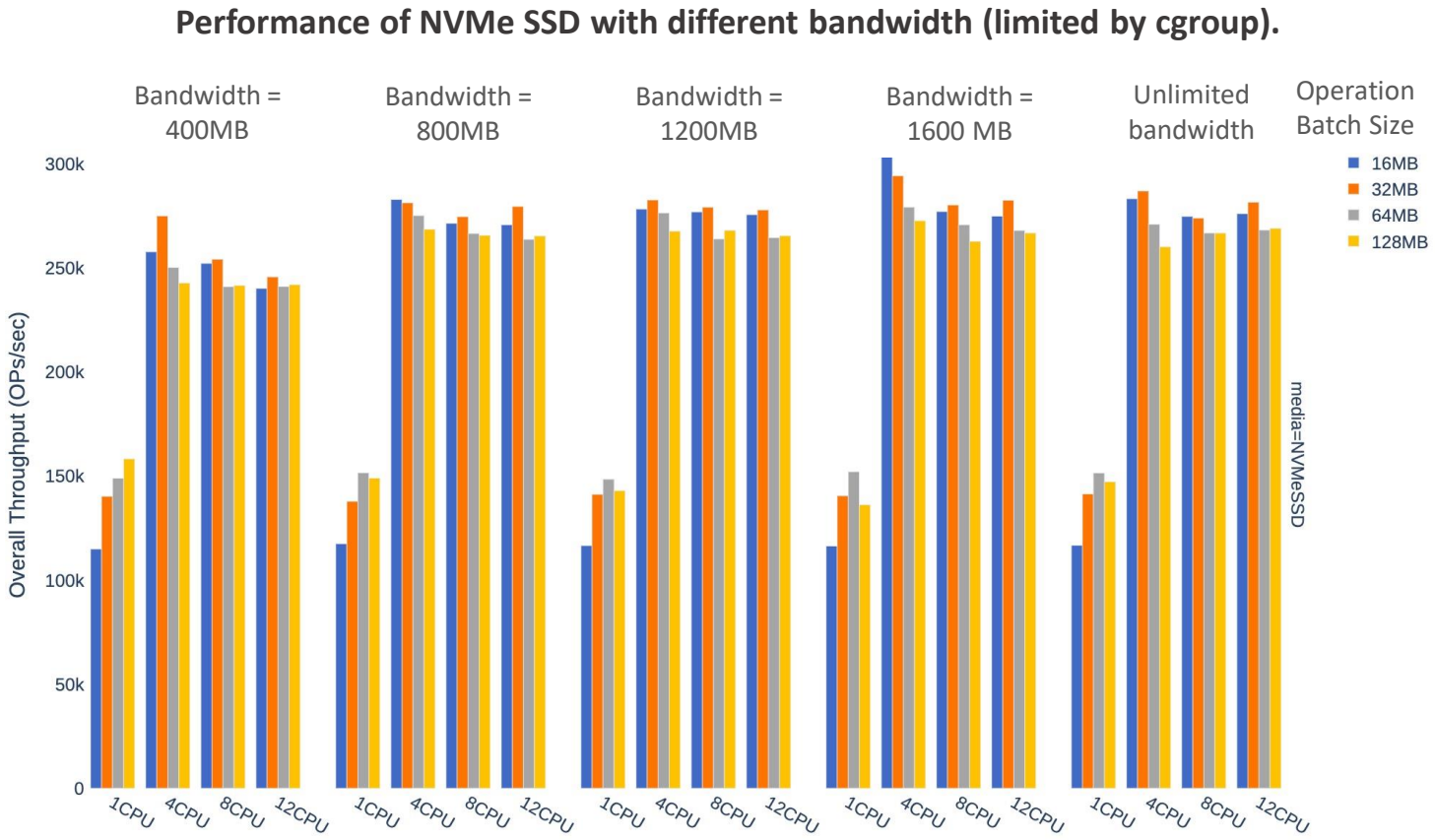
- High bandwidth

Weaknesses

- Larger batch size decreases the performance

Design Opportunities

- Quicksand effect: quicker devices make the data sink too quickly and decreases the performance.
- Improve the pipeline of compaction works



Performance Traits of PMM

Strength

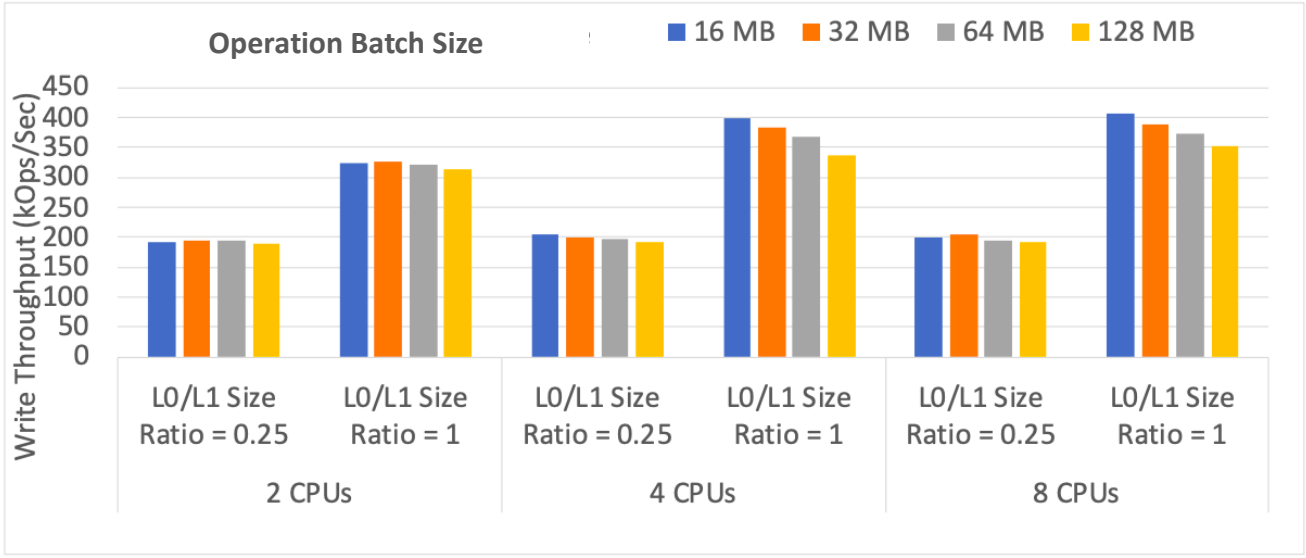
- Stable parallel performance

Weaknesses

- More sensitive to the slow L0 compactions, which can be solved by changing the size ratio between L0 and L1 files.

Design Opportunities

- More flexible data structures
- Can be used directly as a memory expansion
- Non-volatile, free of consistency overhead such as WAL



Size Ratio here means the (total size of L0 files) / (total size of L1 files), controlled by compaction scheduling parameters

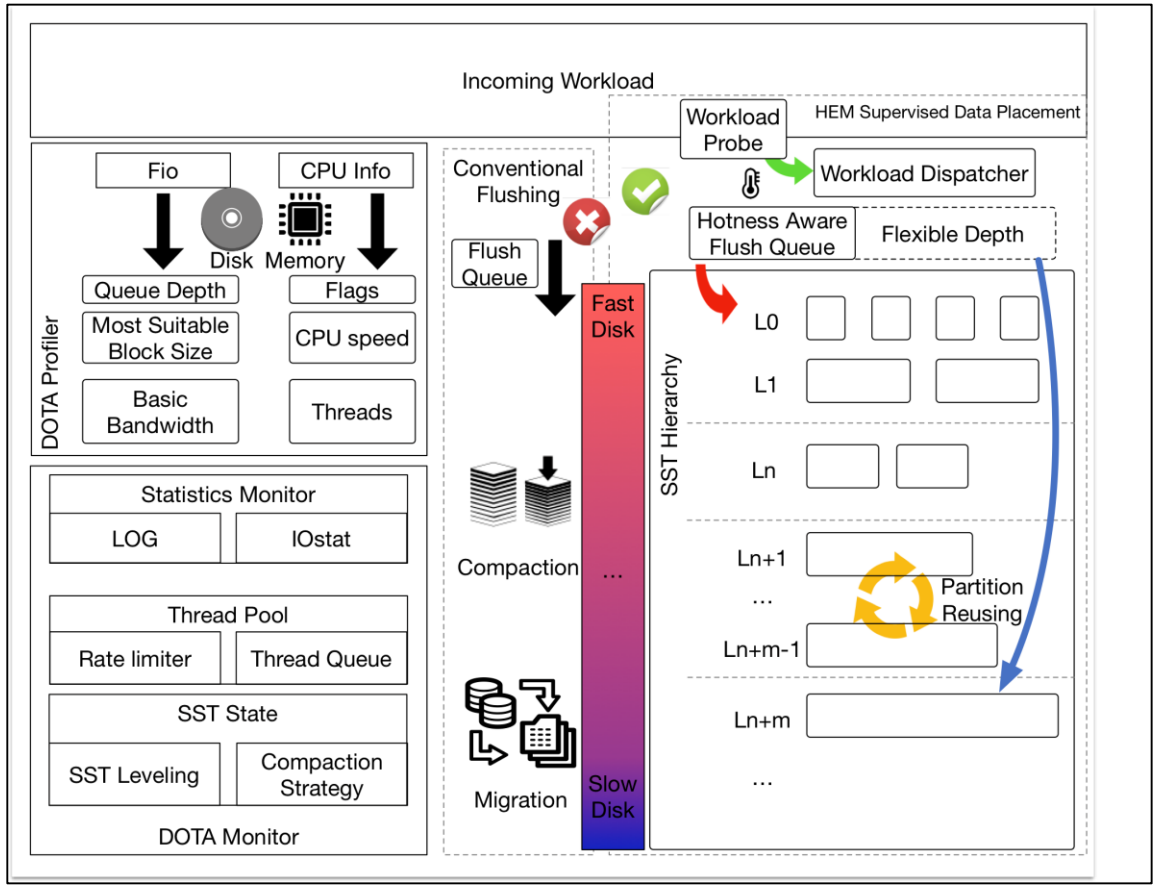
DOTA: Device Oriented Tuning Advisor

Challenges

- Workload adapting
- Online Tuning
- Global resource management
- Environment detecting and monitoring

Solutions

- Online modeling
- Data placement and migration
- Thread pool and resource allocation
- Amplification reduction and data reuse



THANKS FOR WATCHING

Email Address:
jinghuayu2-c@my.cityu.edu.hk

Github link:
https://github.com/supermt/utils_for_lsm.git