SOLUTION BRIEF

Accelerating RocksDB-based Databases with Pliops Extreme Data Processor

Introduction

Cloud workloads and data continue to grow at accelerating rates, as does the demand for better performance, latency and overall efficiency. Unfortunately, computing performance is not keeping pace, while solid-state drives (SSDs) are getting cheaper, denser and faster—especially NVMe SSDs. Today's compute and storage architecture is unable to take full advantage of SSDs, making it more challenging to keep up with growing demands. Nowhere is this more evident with applications like MySQL, Ceph and Spark using RocksDB, an open-source, key-value (KV) storage engine optimized for fast, low latency SSDs. CPU limitations due to higher computational loads and excessive storage I/O consumes valuable compute and storage resources. Without a better technology solution, enterprises have no choice but to deploy more servers and storage—ballooning capital and operating expenses. The Pliops Extreme Data Processor (XDP) is a breakthrough hardware accelerated key-value storage engine that addresses the limitation of a software-only approach and delivers superior database performance and infrastructure utilization.

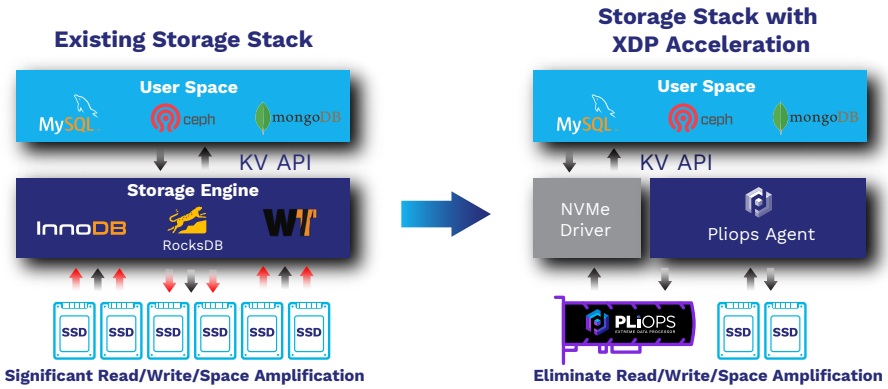

Figure 1: Architecture diagram

The primary performance bottleneck for RocksDB is the write and read amplification that stem from the Log Structure Merge-Tree (LSM) data structure. Each key update in RocksDB leads to many cascading writes in the LSM tree resulting in high write amplification. Sustaining writes to RocksDB requires dedicating a significant portion of I/O bandwidth to perform compactions concurrently. To read a key from the LSM tree, multiple levels need to be verified during read operations resulting in much higher read tail latencies. The compaction process reads several files into memory, sorts them and writes them back at the expense of storage and CPU resources.

Highlights

- 26x TPS gain over RocksDB for mixed read/write workloads
- Improved tail latencies (four nines) up to 740x for writes and 11x read operations
- Enhances flash storage endurance through 6x lower write amplification
- Always-on compression with zero tradeoff for system overhead compared to RocksDB

Features & Benefits

- Simultaneous block and key-value interface support for database and NoSQL applications
- Easy to install and operate KV functions with embedded KV library
- Hardware offloaded data compression and decompression
- Highly resilient: RAID 5-like protection with performance 2x greater than RAID 0 even during rebuilds
- Consistent performance gains by abstracting access to different flash storage devices (QLC, ZNS, TLC and Optane)

Pliops KV Performance Advantages

The primary performance bottleneck for RocksDB is the write and read amplification that stem from the Log Structure Merge-Tree (LSM) data management. Each key update in RocksDB leads to many cascading writes in the LSM tree resulting in high write amplification. Sustaining writes to RocksDB require dedicating a significant portion of I/O bandwidth to perform compactions concurrently. To read a key from the LSM tree, multiple levels need to be verified during read operations resulting in much higher read tail latencies. The compaction process reads several files into memory, sorts them and writes them back at the expense of additional CPU resources.

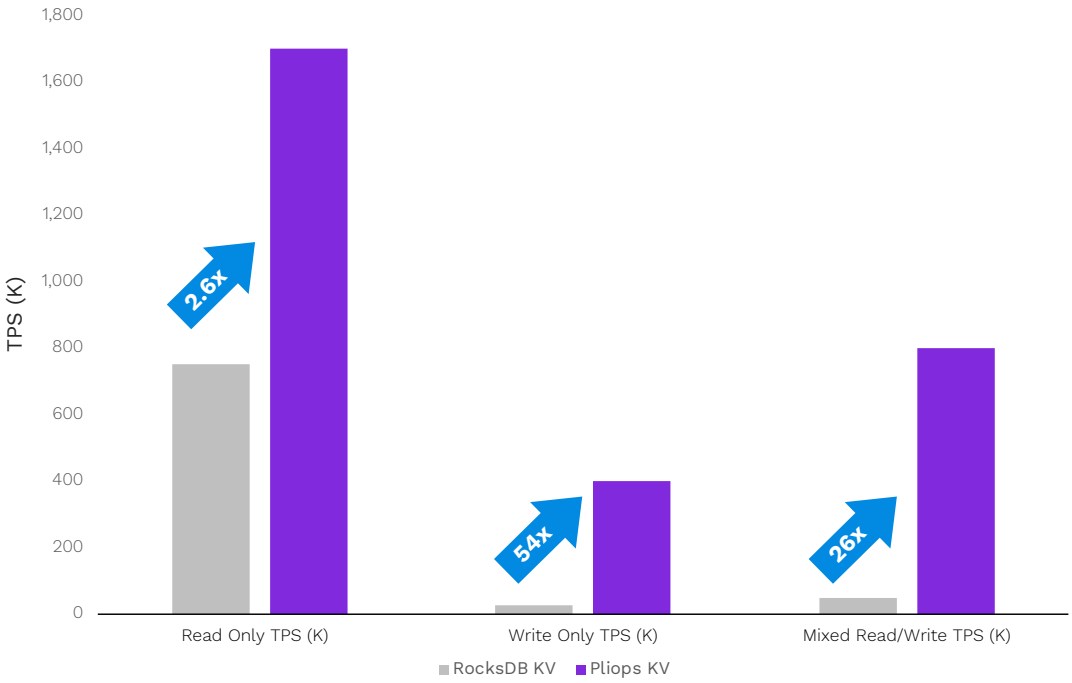

Figure 2: Performance results

Pliops XDP's native key-value-based architecture significantly reduces read and write amplification to their theoretical minimums using innovative data structures and algorithms that optimize write and read operations of key-value data. Finally, the XDP API eliminates the traditional host storage stack overhead, providing applications with direct access to highly optimized hardware that is easy and seamless to deploy in new or existing server deployments.

Technology	Specifications / Version
Processor	Intel Xeon Gold 5218R 2.1Ghz
Memory	512GB
Storage	4.4TB Samsung PM983 SSDs
RocksDB Version	6.2.fb

Table 1: Sytem configuration

Performance Metrics	Read-Only TPS (K)	Write-Only TPS (K)	Mixed Read/Write TPS (K)
RocksDB KV	611	10	35
Pliops KV	1,596	540	918
Performance Gain	2.6x	54x	26x

Table 2: Performance results

Latency Metrics	Read Tail Latency (99.99%) [ms]	Write Tail latency (99.99%) [ms]
RocksDB KV	39	206
Pliops KV	3.6	0.279
Latency Reduction	11x	740x

Table 3: Performance results

The superior performance of Pliops XDP over RocksDB is validated using an industry standard benchmark tool, db_bench. Both XDP and RocksDB are set up on two separate servers with identical configurations. To represent a real-world scenario, the object sizes of the KV store varied from 8 bytes to 10KB. The workloads patterns include DB load of 1 billion keys: loading a new KV DB in random key order, overwriting the entire key space, random read from 1 to 64 readers, and mixed read/write (70/30) workload. The results are shown in **Table 2** and **3** respectively.

Pliops XDP provides an incredible 26x increase in transaction per second (TPS) over RocksDB for mixed workloads, 54x and 2.6x improvements for write-intensive and read-intensive workloads respectively as shown in **Figure 2**. Further, XDP dramatically reduces latencies up to 740x for write and 11x for read operations, as represented in **Figure 3**. This is important because the performance and latency characteristics of a storage engine determine the ceiling limits on database application performance. The test results clearly demonstrate the benefit of both higher performance and ultra-low latency advantage of the Pliops XDP over RocksDB for database applications.

Additionally, read amplification is reduced by a factor of 2.8x and write amplification by 6x as shown in **Figure 4**. The benefit of reduced write amplification enables improved performance and increases the endurance of flash storage media devices. Pliops XDP is also tested and verified to obtain a consistent performance gain on several leading server and flash storage providers in the industry. XDP supports all SSD and flash technologies, including TLC, QLC, ZNS and Optane. Red Hat, Oracle Linux, Cent OS, Ubuntu operating systems also support XDP.

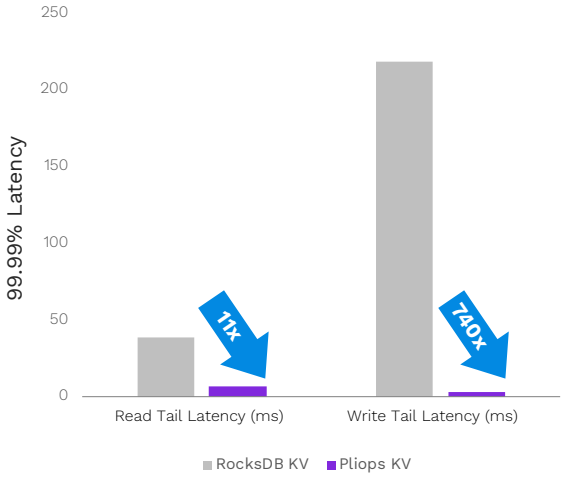

Figure 3: Four nines (99.99%) of latency chart

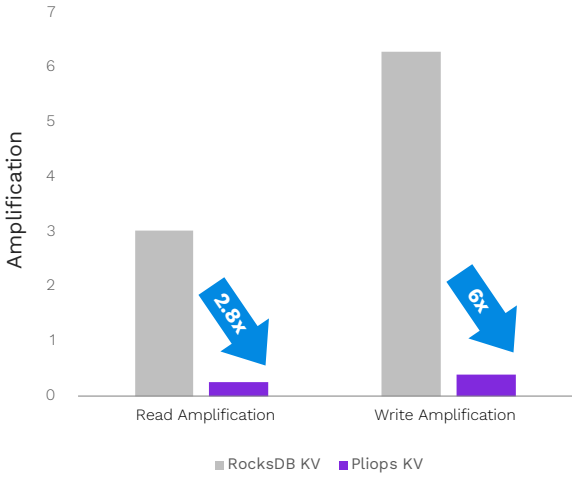

Figure 4: Write and read amplification results

Data protection at speed is essential for high performance databases and applications. XDP's innovative drive fail protection technology delivers RAID 5/6-like protection with performance 2x greater than RAID 0 even during rebuilds while delivering full storage utilization for user data. This enables increased uptime and efficiency of local, distributed and clustered deployment environments.

Conclusion

The Pliops Extreme Data Processor is breakthrough technology designed to deliver superior performance, improved availability, with always-on compression for both key-value and block storage-based applications. It eliminates the inherent design limitations, complexities, and performance drawbacks of RocksDB with a superior hardware accelerated architecture. Test results using db_bench demonstrates XDP outshines RocksDB performance and reduces latencies across all workload patterns, thus increasing quality of service for business applications. The performance boost, latency reduction, improved availability, and KV API functions provide enterprises total confidence to deploy the Pliops XDP for business-critical applications.

About Pliops

Pliops multiplies the effectiveness of organizations' infrastructure investments by exponentially increasing datacenter performance, reliability, capacity, and efficiency. Founded in 2017 and named as one of the 10 hottest semiconductor startups by CRN in 2020 and 2021. Pliops global investors include NVIDIA, Intel Capital, SoftBank, Western Digital, KDT, and Xilinx. **Learn more at www.pliops.com**.