The Ultimate Key Value Accelerator

Why Pliops XDP-Rocks is pivotal for Kvrocks Applications

Executive Summary

Kvrocks is an open-source distributed key-value NoSQL database for storing and processing large datasets. It is powering many business-critical applications at Baidu, Ctrip, Meitu, RGYUN, U-NEXT, and BaiShan enterprises. It was developed to provide an open-source alternative to Redis to process the datasets on disk. It uses RocksDB as a storage engine and is compatible with the Redis protocol. Unlike Redis, Kvrocks serves as a persistent key-value store reducing the memory cost in dealing with large datasets. As this dataset grows to petabyte scale, performance and latency prove to be challenging to meet customer SLAs due to inherent limitations of the RocksDB storage engine. To scale Kvrocks deployments, infrastructure optimization and efficiency are needed to not only reduce the data footprint cost but to also drive a delightful experience for customers.
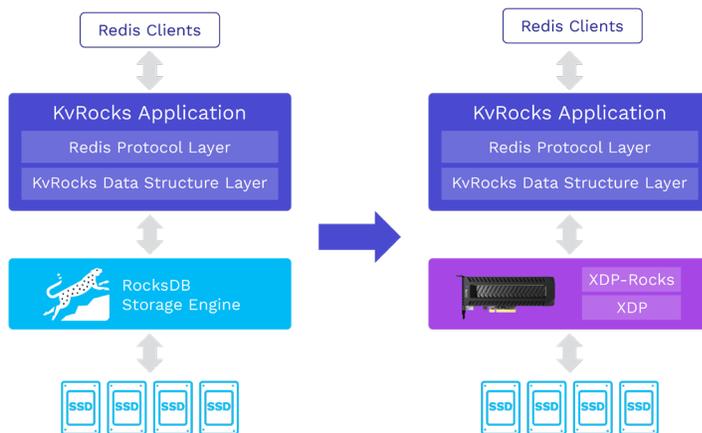

Figure 1: RocksDB to XDP-Rocks transition for Kvrocks Applications

RocksDB, which supports Kvrocks as a storage engine encounters significant performance impact during the data compaction process. The compaction process results in write amplification, and space amplification which effectively impacts SSD lifespan and its capacity utilization. Compaction also results in a significant drop in performance for data writes, scan, and lookup operations of the Kvrocks applications. Based on the data access patterns and data ingestion rates business applications may suffer service level objectives and customer satisfaction due to the compaction process.

Key Highlights

Pliops XDP-Rocks delivers unique & compelling benefits for Kvrocks Application to improve enterprise performance and efficiency.

Up to **11x** higher performance for write operations

Up to **51x** tail latency reduction for write operations

Up to **10x** higher performance for mixed workload performance

Up to **2.5x** higher SSD Endurance

Accelerating Kvrocks with Pliops XDP-Rocks

Pliops XDP-Rocks is a Software library that provides RocksDB-compatible API for integrating with the Kvrocks application. XDP-Rocks provides a rich set of data storage engine features from basic put/get/delete operations to advanced scan, and range operators for large dataset retrievals. It also incorporates row cache, checkpoint, and compaction filter functionality to improve the performance of read and write-intensive Kvrocks applications. XDP-Rocks eliminates the inherent architecture limitations of the RocksDB compaction process.

Thus, it helps to increase the write throughput with reduced write amplification. This document summarizes evaluation results comparing Kvrocks vanilla (sw v2.0.6) vs XDP-Rocks-based Kvrocks.
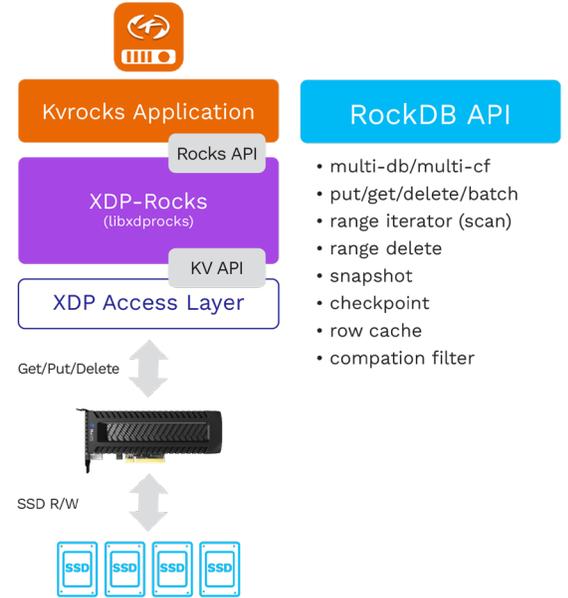

Figure 2: XDP-Rocks Integration with Kvrocks

Kvrocks Memtier Performance Testing

Memtier benchmark was employed to evaluate the performance benefits of XDP-Rocks with Kvrocks (Kvxdprocks) over RocksDB with Kvrocks. The benchmark results compare overwrite, read-only, and mixed workload operations with varying number of threads, measuring throughput and tail latencies.

We selected 512B and 4KB object sizes of 4 the Terabyte dataset to serve various customer application needs.

For both 512Byte and 4K object size datasets, Memtier benchmark was executed with the fill, write-only (overwrite), read-only, and mixed (50% read, 50% write) sequence of operations. The fill stage workload is parallel-sequential running 16 threads per database and each thread loads non-overlapping key ranges. In the overwrite stage 1 thread per database runs several operations that are equal to the number

of objects in the database which is 1Billion for 512Byte and 125Million for 4KB. All other steps are executed for a 30-minute duration with varying threads. Keys are sampled from a uniform random distribution.

The system saturation point is identified by increasing the number of threads in each of these workloads. Beyond this saturation point, latency increases significantly with almost no gain in throughput. This is known as the knee-point in the graph.

We compare Kvxdprocks knee-point performance vs Kvrocks knee-point performance. In the caption of the figures below these are referred to as the saturation points. Further, we compare the best throughput each system achieves with bounded tail latency p999<10ms. reserving the fifth SSD as a hot spare for rebuild use only.

Object Size	Total objects per Database	Number of Databases	Total Dataset
512 Bytes	1 Billion	8	4 TeraBytes
4K Bytes	125 Billion	8	4 TeraBytes

RocksDB vs XDP-Rocks for 100% Write Intensive Operations

For a 512bytes object, size Write performance throughput with XDP-Rocks over RocksDB is increased by 1.4x during the fill stage (Data loading) and 3.4x during Overwrite operations as shown in **Chart1**. For 4K object size, the performance benefit is even higher during the Fill stage (Data loading) it increases to 4.8X and overwrites operations performance increased to 10.6x.

Latency reduction is another significant benefit, and it was realized for the entire spectrum of latencies that is from p50(average), P99 (tail) latency to P99.9 (tail latency). Overwrite P99.9 latency has reduced from 26.88 milliseconds to 0.53 milliseconds that's roughly 51X latency reduction for 4K object size and 512 bytes p99.9 latency reduced by 10.3X. This benefit is extremely important for high data ingestion rates applications looking to significantly reduce 4 9's latencies and thereby improve customer satisfaction.
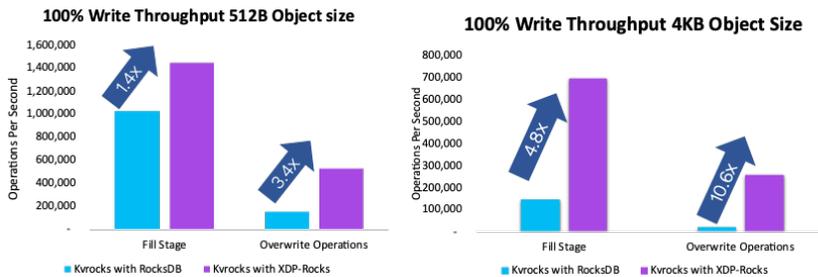

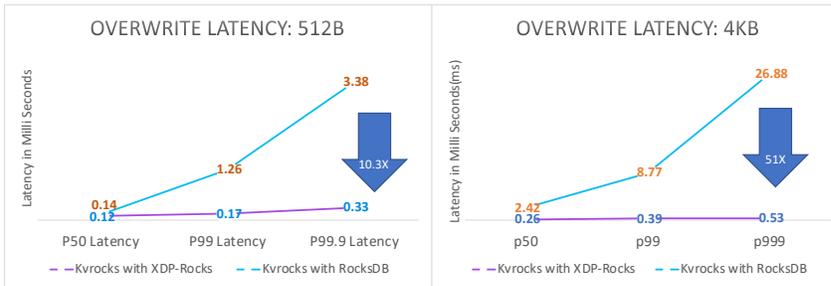*Chart1: XDP-Rocks Performance benefits for Write Operations*

Chart2: XDP-Rocks Performance benefits for Write Operations

RocksDB vs XDP-Rocks for 100% Read Intensive Operations

For a 512bytes object, size Random read performance throughput with XDP-Rocks over RocksDB is increased by 2.5X while reducing the tail latency from 15 to 44X. Another key observation from this experiment XDP- Rocks ensures the tail latency is under 10 milliseconds irrespective of the number of concurrent threads accessing the dataset Sametime RocksDB latencies shoot up to 160 Milli seconds. This benefit is crucial for session caching applications to deliver consistent latency and Quality of service for customer-facing applications.
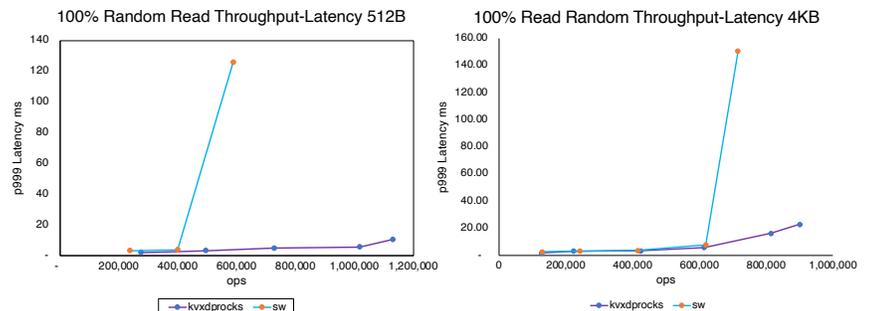

Chart3: XDP-Rocks vs RocksDB for 100% Random Read

RocksDB vs XDP-Rocks for Mixed Workload Operations

In a mixed workload scenario, As shown in chart 4 Kvrocks running with RocksDB achieves no performance scaling with the increased number of user threads. The performance remains flat at around 50K Ops. However, the same tests with XDP-Rocks and Kvrocks integration deliver excellent performance scaling with 100K ops for single user threads to 650K ops for 64 user threads.
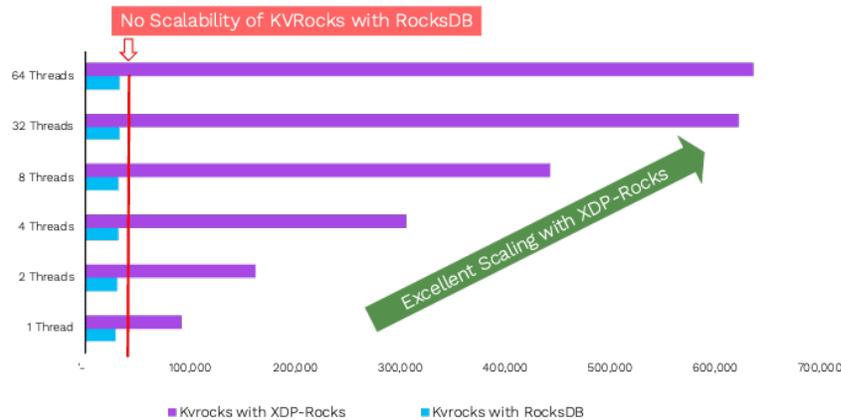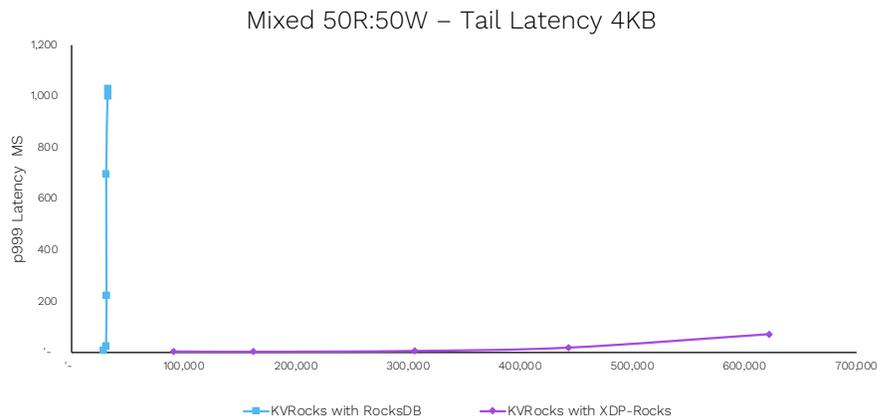

Chart4: XDP-Rocks vs RocksDB for Mixed Workload

Mixed 50R:50W – Tail Latency 4KB

Conclusion

The Pliops XDP-Rocks is pivotal to KV-Rocks applications to deliver massive performance boost and reduce orders of magnitude latency reduction to dramatically improve service level objectives and customer experience. These benefits are readily applicable to Redis and Memcached applications and they will be discussed in a separate solution brief. XDP-Rocks enables enterprises to accomplish infrastructure optimization and drive efficiencies to significantly reduce capital and operational expenditure. With easier XDP-Rocks, integration enterprises can efficiently address the growing business demands of application scaling and address data center constraints in dealing with petabyte-scale datasets.

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**.