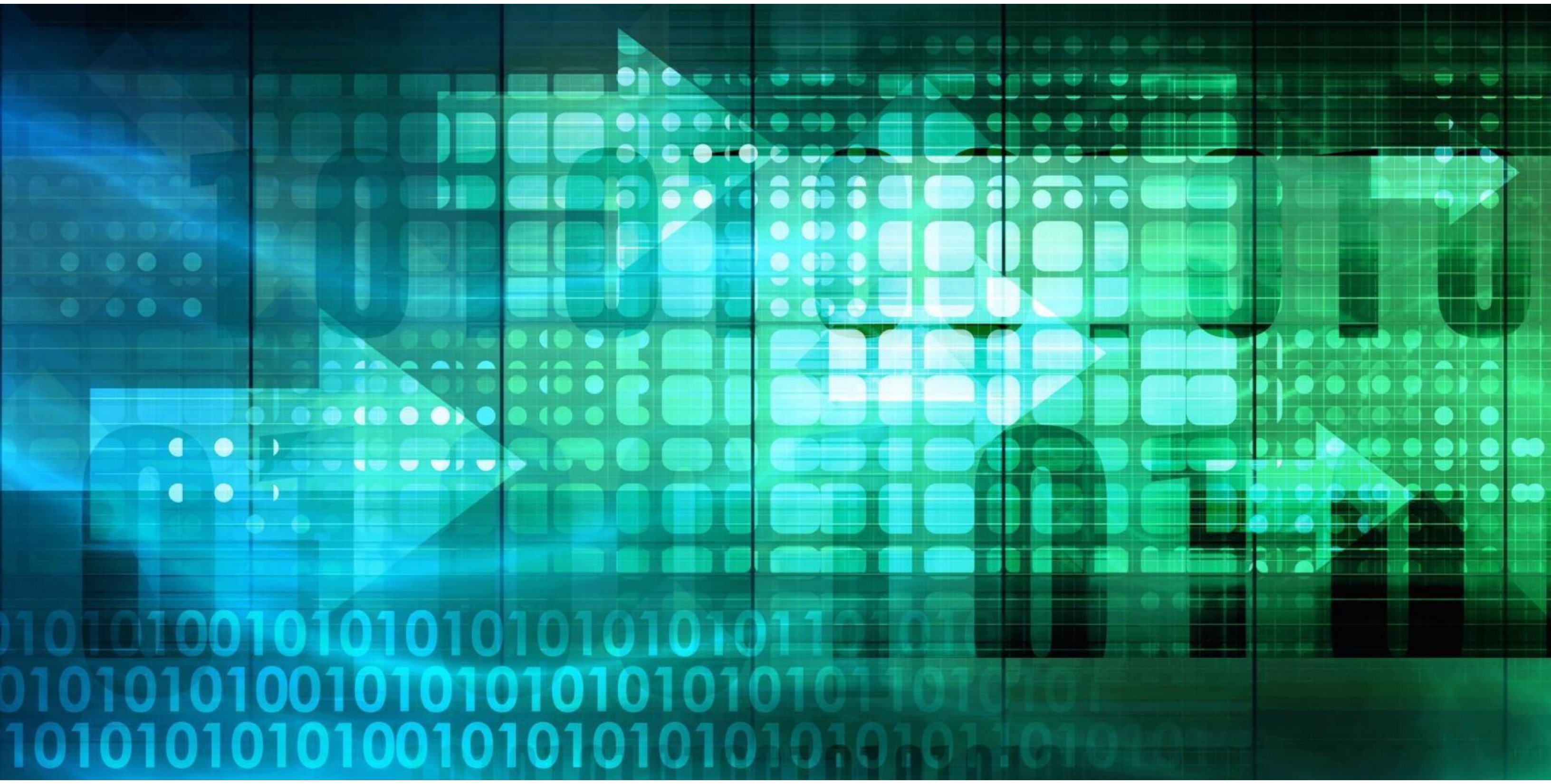
PLIOPS

WHITE PAPER



Pliops Storage Processor (PSP) Overview

HIGHLIGHTS

- Compute performance growth predicted by Moore's Law is slowing to a crawl[1] at a time that storage and network performance growth is accelerating, resulting in server sprawl and increased data center costs.
- PSP implements a completely new architecture to accelerate storage-intensive applications—taking advantage of dedicated hardware processing to accelerate performance—that improves response times and reduces CPU, networking and storage demands to lower cost and increase scaling.

The Pliops Storage Processor offers breakthrough economics and performance, as well as data reduction for a wide range of data center applications. By redefining the storage primitive to be directly accessible by the application, Pliops brings the system back into balance across compute, network and storage resources, enabling companies of all types to meet the growing workload demands posed by the growth in cloud databases, analytics and new growth drivers such as AI/ML, 5G and IoT.

This white paper describes the Pliops Storage Processor and how it integrates with key-value store (KV store) environments.

BACKGROUND

The End of the Moore's Law Era

Moore's Law faces challenges and there is no foreseeable path to resume the easy performance scaling of past years. Computing performance has slowed from its prime of doubling every 2 years and is now forecasted to double every 20 years, according to UC Berkeley's David Patterson. The impact on data centers from the end of Moore's Law is already being felt as businesses experience server sprawl driving capital and operating expenses ever higher—especially for cloud-scale and rack-scale environments.

Leading organizations must begin planning for an era where workload-specific accelerators are required to continue scaling their data centers. These accelerators substitute for faster server processors that will remain unavailable, and free general-purpose CPUs to focus on functions only they can perform.

An Era of Rapid Storage and Networking Advances

With NVMe-based solid-state drives (SSDs) becoming mainstream, storage that is 1,000 times faster than hard disk drives (HDDs) have created unique opportunities for evolving data center architectures.

Networks deployments with speeds as fast as 400Gbps (400G) are being planned now and new highly efficient network protocols like NVMe over Fabrics (NVMe-oF) will enable expanded use of ultra-fast storage. Unfortunately, databases and applications and the infrastructure to support them have not kept pace with the growth of storage and networking performance.

New Architecture Required

Key-value (KV)-based storage engines are used to manage data persistence and indexing tasks common to transactional databases, analytics applications and software-defined storage (SDS).

The Pliops Storage Processor uniquely accelerates these storage engines and the applications they support. The first-generation PSP will be deployed as a low-profile PCIe add-in card including the software required to utilize it seamlessly with the most demanding databases and applications.

When PSP is compared to RocksDB software, one of the most advanced and popular open source storage engines available, the Pliops Storage Processor supports 13x more read/write performance due to its accelerated architecture. By also offloading workloads from server CPUs, the number of server processor cores being used for a given workload is reduced to less than 2% versus RocksDB software.

KV STORAGE ENGINE PROBLEMS

One of the typical byproducts of KV store operations is an increase in read, write and space amplification. As the term suggests, each write from an application to the storage engine is amplified by internal storing and sorting mechanisms.

Amplifications levels of 20x, 40x, and even 100x are common and this drives huge demands on networks and SSDs. Infrastructure teams solve this problem today at great expense, and with Moore's Law slowing to a crawl, that expense is only growing.

The specific challenges created are:

- **Space Amplification**
  Databases can store variable-length data but SSDs store data as fixed-length blocks that are much larger than the natural storage unit of most applications. The methods used by the most common applications to store and index data require data structures to not be full, resulting in wasted storage space. Applications can improve performance by expanding the utilization of storage space, with a great cost of additional SSD storage.

- **Read Amplification**
  Databases retrieve data in complete blocks from storage systems, even when a smaller portion is needed, so small read requests can become much large ones. Many flash-based applications amplify reads by a factor greater than 100.

- **Write Amplification**
  Similar to the phenomenon on SSDs, storage engines either perform garbage collection or must write complete blocks when the data to be updated is much smaller than a block. This causes blocks of data to be written, erased and rewritten multiple times. In addition, the processes used to keep indexes sorted initiates multiple writes for each application IO. This consumes PCIe bandwidth for direct-attached SSDs, or network bandwidth if SSDs are disaggregated. It also drives storage costs higher as storage architects limit the utilization of the drive to provide sufficient SSD endurance. In addition, high levels of writes interfere with reads within an SSD, creating high read latencies and poor QoS.

- **CPU Utilization**
  The previous problems can be addressed with more aggressive garbage collection and rebalancing techniques managed by the application. However, application architects avoid these techniques as they would add further stress to CPUs and memory. Ultimately, there is not enough processing power to deliver optimal levels of space, read and write amplification.

Overcoming these challenges begins with database architects who make decisions about server and storage hardware configurations. Database administrators then test and tune database and application software to get the best performance the system can support.

Database and application performance becomes limited by system hardware—and modern IT infrastructure has become imbalanced with storage (and networking) performance advancing faster than processor and memory performance—so a better solution than adding processors, memory and drives is required.

That solution is the Pliops Storage Processor.

THE PLIOPS SOLUTION

Game Changing Performance

The Pliops architecture utilizes a new data structure for database-related storage operations—indexing, searching, sorting, merging, etc.—and accelerates them with purpose-built hardware to deliver profound performance gains, while not requiring any software changes from user applications

One way to measure Pliops Storage Processor performance improvements is to compare PSP to the RocksDB storage engine.

PSP Key-Value Performance	Storage IOPS Increase	CPU Workload Reduction
Load	23x	97%
Overwrite	5.1x	99%
Read/Write	13x	99%
Get	2.4x	82%

KV Performance: PSP Hardware Versus RocksDB Software

The previous table captures the benefits of PSP and shows the effects of its reduction of read and write amplification. This enables a single server with a PSP included to support the same performance of nearly 1,000 processors while significantly reducing the cost of SSDs. This improvement in storage engine performance can benefit many of today's most common and challenging applications and workloads.

Unrivaled Response Time

The latency (or response time) benefits of PSP hardware are clear when compared to RocksDB. For example, PSP reduces load latency by three orders of magnitude and improves read/write latency by two orders of magnitude.

99.99% Latency (1 microsecond)	RocksDB	PSP	Benefit
Load	29,040	21	1,383x
Overwrite	39,150	3,237	12x
Read/Write	121,342	1,126	108x
Get	26,320	1,262	21x

KV Latency: PSP Hardware Versus RocksDB Software

The previous table includes data captured during benchmark testing with the benchmarking tool db_bench. The benchmarking system was a server with one Intel 4.3 GHz 6-core CPU, one 1TB NVMe SSD, and 32 GB of memory.

Testing measured queries per second (QPS) and tail latency, with all tests performed after 2 hours of garbage collection for SSD pre-conditioning.

The workloads tested included:

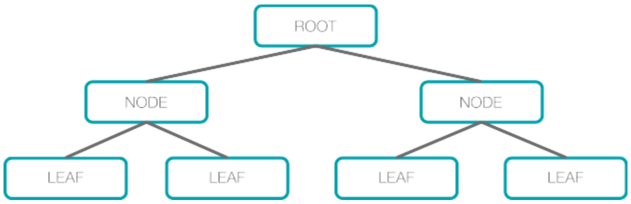
- **Load**
  Loading a new database in random key order as is typical during rebuilding or re-sharding operations.
- **Overwrite**
  Overwriting the key, such as when updating an entry.
- **Read/Write**
  Mixed read/write workload using 1 writer concurrently with 32 readers.
- **Get**
  Random read workload using 32 readers.

WHY PSP WORKS

In general, Pliops Storage Processors works by eliminating the inherent inefficiencies present in databases and applications using those databases. When using software-based storage engines such as RocksDB, these inefficiencies cause significant read, write and space amplification that consumes resources and slows performance.

Storage Engines Overview

Many storage engines organize data and indexes using B Trees, a structure that simulates a hierarchy using a root linked to multiple nodes that are subsequently linked to multiple leaves.



Key-value Store: B Tree Data Structure

One example is InnoDB, the default storage engine of MySQL. By design, B trees support high capacity and large scaling by moving fixed-size data blocks between system memory and storage as required.

Some of the most useful features of B Trees—the abilities to randomly read/write data and update data in-place—lead to serious inefficiencies. For example, it is highly unlikely, the larger the disk/ratio the less likely, that data requested during a random read will be found in memory so the data must be retrieved from storage instead, which amplifies a small IO request to a large one matching the data block size. This significantly amplifies reads.

B Tree Optimization

If storage engines chose to run at maximum efficiency with each leaf 100% full, then each time it needs to accept a new IO, it must open up the entire leaf, split, perform garbage collection, and rebalance across the tree. This would generate enormous strain on CPUs and storage and would be completely infeasible.

Therefore, the system does not allow each Leaf to get full. This means that the system is built to waste storage space to improve performance. The emptier the leaf, the higher the performance. Many architects will try to extract more performance by keeping the Leaf structure emptier, which consumes more storage space. This illustrates the inherent tradeoff for storage engines: space efficiency comes at the cost of performance degradation and CPU and memory overheads.

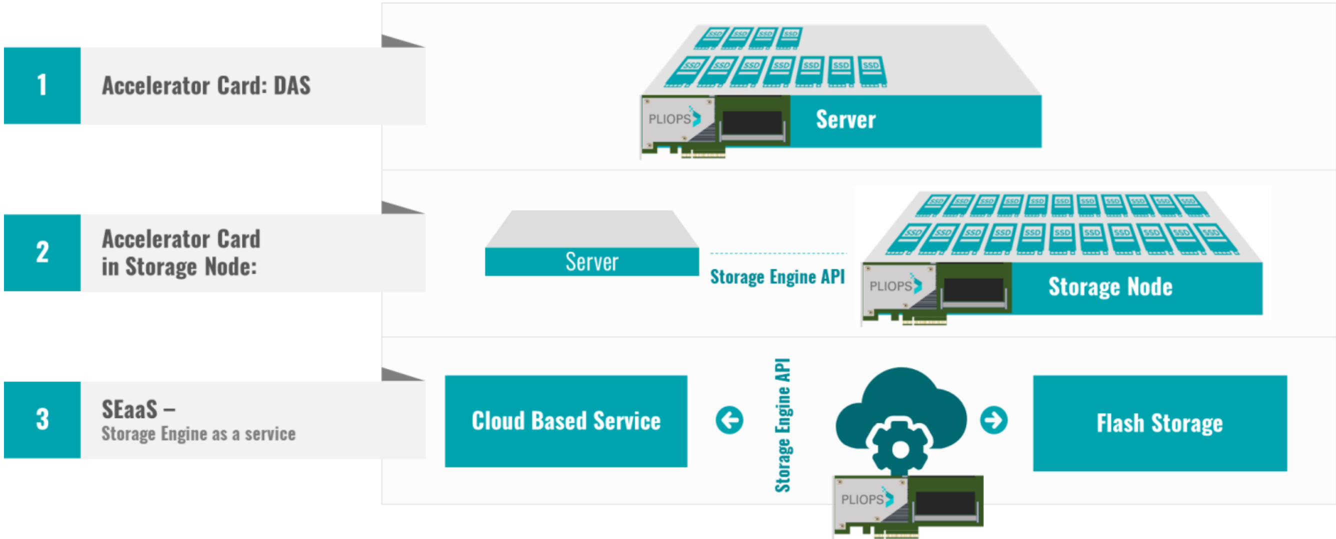
In B Tree architectures, examples of space amplification of 2x to 7x, varying based on data entropy, are common and multiple architectures using this very expense performance-enhancing technique have been published by leading technology providers.

The PSP is built from the ground up around a flexible, dynamic storage structure which means it can achieve the incredible performance levels noted prior, all while keeping space amplification close to the optimal value.

In fact, by efficiently storing these fundamental data structures, Pliops Storage Processor can save space—up to 50%—over host- and software-based compression techniques that cannot overcome the inefficiencies of the leaf structure.

PSP is Easy and Flexible to Deploy

First-generation Pliops Storage Processor will be packaged as a cloud-based service and a low-profile PCIe card. That allows flexibility of deployment to support traditional and modern data center infrastructure and goals.



PSP Deployment: Options for Three Architectures

The Pliops Storage Processor offers three deployment options to support popular data center architectures.

1. **Compute Node Acceleration**
   Adds a PSP card into servers focused on database processing workloads.

2. **Storage Node Acceleration**
   Adds a PSP card into storage serving storage engine workloads.

3. **Cloud Service Acceleration**
   Uses PSP cards as a service for cloud-based key-value workloads.

Regardless of the deployment options selected, PSP presents a Storage Engine API that enables directly maximizing storage performance (and indirectly increasing database and application performance) as well as actively minimizing read, write and space amplification.

ABOUT PLIOPS

Founded in 2017, Pliops is a technology innovator focused on making data centers run faster and more efficiently. Its technology addresses skyrocketing data volumes and solves the slowing compute performance problem. The company's storage processor is built upon a groundbreaking patent-pending approach that accelerates storage functions. Focused on creating the next wave of the accelerated data center, Pliops' storage processor enables cloud and enterprise customers to access data up to 100 times faster – using just a fraction of the computational load and power consumption. With Pliops technology, databases, analytics, and other data-intensive applications are able to reach their full potential. Investors include Softbank Ventures Asia, Intel Capital, Western Digital, Mellanox and Xilinx.

---

1. https://californiaconsultants.org/wp-content/uploads/2018/04/CNSV-1806-Patterson.pdf