ESG Lab Spotlight

PrimaryIO Application Performance Acceleration

Date: July 2015 **Author:** Tony Palmer, Senior Lab Analyst

Abstract: *PrimaryIO Application Performance Acceleration (APA) is designed to provide hardware-agnostic, application-aware performance acceleration for physical or virtualized servers. APA leverages server-side flash and is engineered to target only high-value application I/O for acceleration. This spotlight summarizes ESG Lab's validation of the performance benefits offered by APA.*

Background

According to ESG research, among the top ten most important IT priorities reported by respondents for 2015 are the increased use of server virtualization, management of data growth, major application deployments or upgrades, and data center consolidation.[1] These responses strongly suggest that IT is facing increasing pressure to improve efficiency while continuing to deliver nonstop application and data access.

Achieving these objectives is difficult with silos of traditional disk storage. Deploying, configuring, managing, and scaling storage for different applications adds complexity, which increases both the time and cost of providing infrastructure services to end-users. Widespread adoption of solid-state storage began with the introduction of flash memory as a disk tier. Whether deployed as a tier or as a cache, solid-state drives (SSDs) are now being used to improve the performance of business applications such as OLTP databases, ERP financials, OLAP business intelligence, and high-performance computing. SSDs are also becoming popular to accelerate the performance of both bare-metal and consolidated virtual server and virtual desktop infrastructures, but the challenge is doing so in a way that is cost-efficient and seamless to existing operations.

PrimaryIO APA

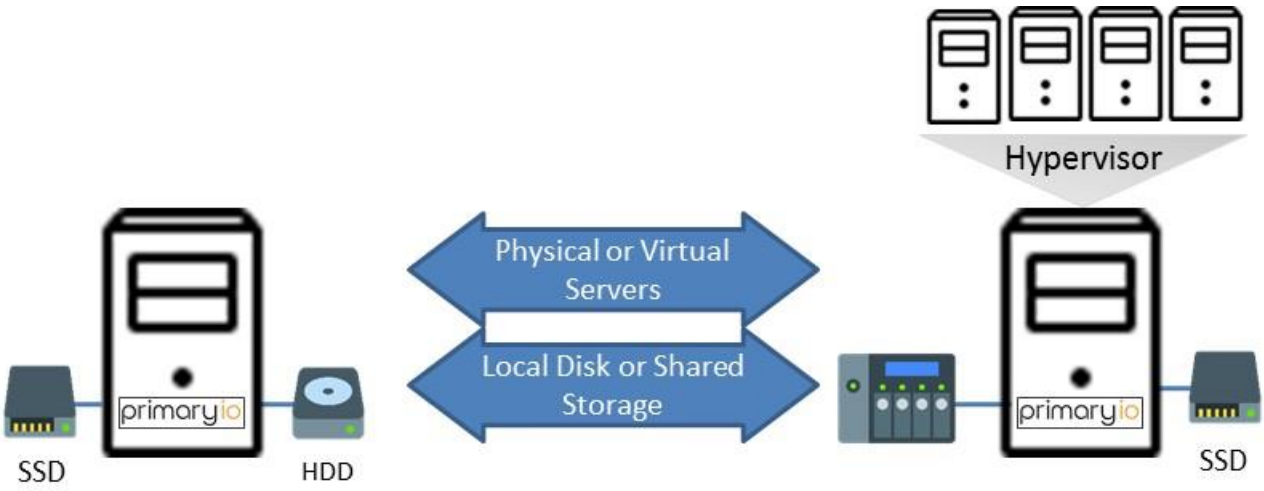
APA is a software-only, hardware-agnostic application acceleration solution designed to leverage server-side flash to optimize application I/O in a targeted manner. The PrimaryIO solution delivers uniform acceleration capabilities across a Windows or Linux bare-metal server or a virtualized environment. Applications can be run within a private data center or on public clouds like AWS, Google Cloud Platform, or Rackspace. The installation and setup are fast and simple and do not require application reconfiguration, which can affect storage workflows.

APA leverages an application-level intelligence component called Application Performance Acceleration (APA) engineered to provide primary data and I/O acceleration for business-critical workloads with easy administration. The APA technology works by intercepting application I/O and determining the value of each application element to identify and accelerate performance of the most important data to an application or business. Database transactions stored in tables, indexes that speed up queries, or the web pages that are most important to a business are just a few examples. Tier-1 databases supported include Oracle, SQL Server, MYSQL, and MongoDB; all can be optimized to run up to ten times faster without any modifications to the application.

PrimaryIO's technology is designed to complement modern, virtualized, and cloud data centers powered by flash. APA runs on physical or virtualized servers, in traditional data centers or on Amazon Elastic Compute Cloud (EC2) instances.

[1] Source: ESG Research Report, *2015 IT Spending Intentions Survey*, February 2015.

Figure 1. PrimaryIO APA



Key capabilities and benefits of PrimaryIO:

- **Application Performance Acceleration** – Support for Oracle, SQL Server, MongoDB, and MySQL that is designed to provide targeted performance acceleration, improved flash longevity, and network utilization. For all other applications, APA supports generic block-level caching.
- **Integrated Workload Analyzer and Provisioning Tool** – Analyzes the workload, and recommends which volumes to accelerate, as well as the amount of flash required in order to achieve performance SLAs.
- **SAN, DAS, and Cloud Storage Support** – For a variety of deployment scenarios.
- **Software Only**– The hardware agnostic architecture enables users to choose from a large number of flash offerings to balance price and performance.
- **Application Transparency** – No changes to application configuration are required since APA supports seamless installation as well as the ability to enable or disable the caching engine non-disruptively.
- **Intelligent Flash Management** – Reduces write amplification to deliver consistent performance and extend the useful lifespan of flash.
- **Lightweight** – APA consumes less than 4% CPU overhead and approximately 40MB of RAM.
- **Easy to Implement and Use** – The system self-optimizes acceleration and the web GUI and CLI are designed to further simplify management. APA performance metrics such as transactions per second (TPS) should be intuitive for any IT administrator and do not require deep systems expertise.

Application Performance Acceleration

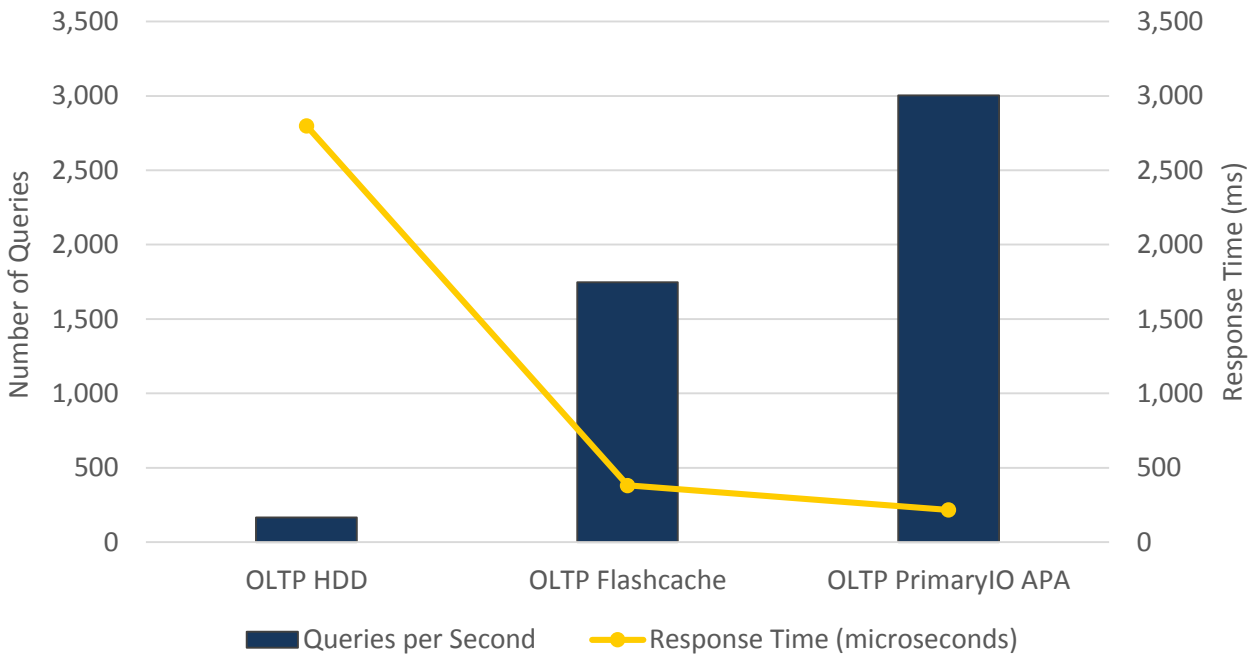
ESG Lab audited PrimaryIO's performance acceleration testing of three real-world workloads: online transaction processing (OLTP), online analytics processing (OLAP), and virtual desktop infrastructure (VDI). Performance tests were run on an industry-standard x64 server with three eight-core CPUs and 4GB of RAM, one Micron P320H SSD, and one 10KRPM HDD installed. The server was running Ubuntu 12.04 LTS inside a KVM virtual machine. The small onboard RAM configuration was deliberate, to ensure the dataset would not fit entirely into RAM. OLTP testing used Sysbench v 0.4.12 against a MySQL database with a 23GB dataset. OLTP testing was executed using both APA, and Flashcache, an open source caching software package.[2] The I/O profile was 100% random, with 70% reads and 30% writes. OLAP Testing was accomplished using the Yahoo Cloud Serving Benchmark (YCSB) tool, running against a MongoDB database with a 17GB dataset. The OLAP workload was characterized by a long insert of log records—100% Sequential, 100% writes—followed by queries against the newly inserted data, 100% random, 100% reads.

[2] https://github.com/facebook/flashcache/

The first step in OLTP testing was to install APA and analyze the application I/O to determine the optimal amount of cache for the application. Installation is scripted for simplicity, completely transparent, and non-disruptive to applications. APA can automatically detect the application and monitor I/O patterns. The default period to monitor is 12 hours. While users can adjust that timing to suit their needs, PrimaryIO best practices recommend allowing the software to analyze I/O for the full 12 hours. When I/O monitoring was complete, the installation script recommended which volumes and application components (e.g., tables, indexes) to accelerate and how much cache would be required to accelerate those volumes. It's important to note that the installation—including monitoring—was executed with Sysbench running against the database. The workload continued to run through the installation and optimization phases, with no disruption or impact. Acceleration was completely automatic.

The results of OLTP testing are shown in Figure 2. The single hard drive was able to support 166 queries per second with an average response time of 2.8 seconds. While Flashcache was able to improve hard drive performance, APA was able to drive 3,003 queries per second—18x the performance of the hard disk—with a response time of just 212ms, nearly double the performance of Flashcache, with about half the response time.

Figure 2. MySQL OLTP Performance Acceleration



Management of SSD write activity is critically important because MLC flash has a limited number of writes it can support during its lifetime. During OLTP testing, ESG Lab observed that APA wrote just about half as much data to SSD as Flashcache.

Next, ESG Lab looked at analytics. Figure 3 shows the load phase. The single hard drive was able to support nearly 8,000 queries per second with an average response time of 4ms. This is not surprising since large block sequential workloads are what hard drives are best at. APA was able to sustain just under 13,000—nearly 2x the performance of the hard disk—with a response time of just 2.4ms, which is nearly identical with the results produced by the SSD. ESG Lab also examined the run phase, where queries are run against the newly loaded data. In this case, APA again provided performance extremely close to that of the SSD, servicing more than 90,000 queries per second at a response time of just 341 microseconds.

Figure 3. MongoDB OLAP Performance Acceleration – Load Phase

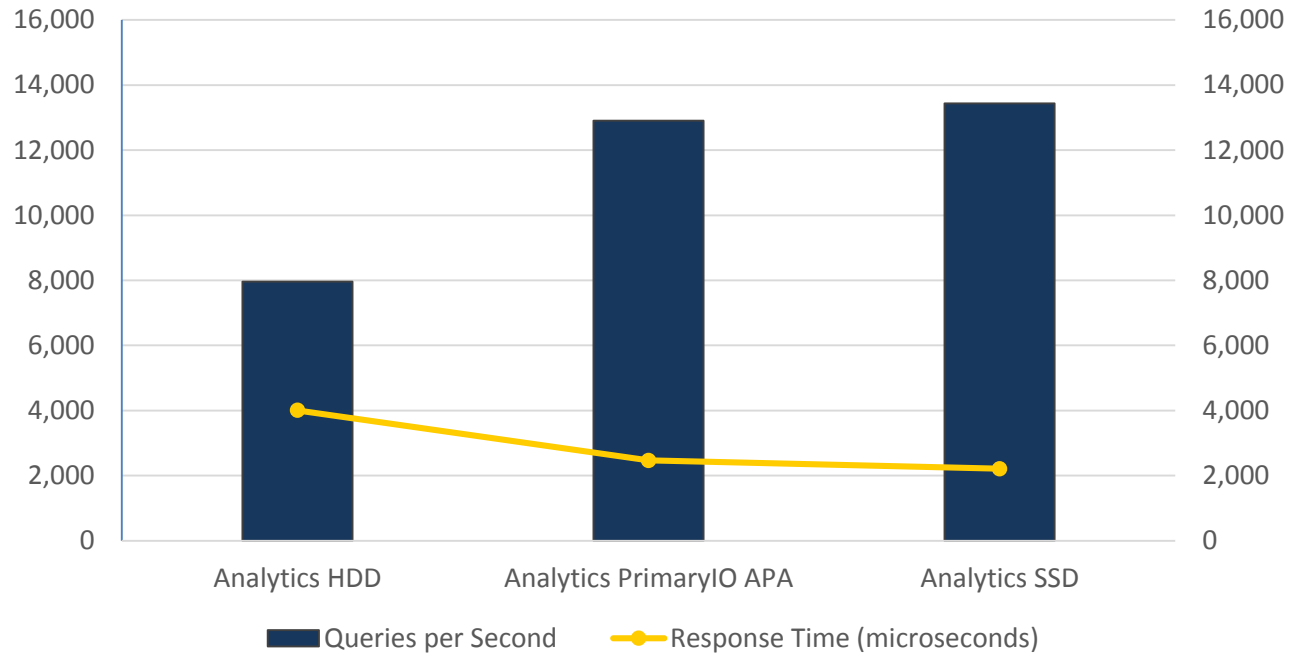


Table 1 shows the detailed results of testing.

Table 1. APA Performance Acceleration Details

OLTP (SQL Server)	Transactions per Second (TPS)	Queries Per Second (QPS)	Response Time (ms)	HDD IOPS	SSD Writes (GB)
HDD Baseline	11	167	2,800	212	-
Flashcache*	84	1,747	382	181	21
APA	147	3,003	217	147	11
Analytics (MongoDB)	**Load Phase QPS**	**Load Phase Response Time (µs)**	**Run Phase QPS**	**Run Phase Response Time (µs)**	
HDD	7,965	4,005	207	153,913	
APA	12,911	2,468	90,256	341	
SSD	13,436	2,212	94,891	323	

* Flashcache required 550 MB of RAM to manage a 100GB SSD, which resulted in some observed swapping during this test.

Finally, ESG Lab examined generic acceleration with APA using a Real Desktop Experience (RDE) VDI environment with 100 thin clients. When thin clients boot for the first time in an RDE environment they write an average of 400-500MB of data to the server. Large numbers of clients booting simultaneously is commonly known as a boot storm, where the large number of random writes causes extreme performance degradation in disk-based environments. Clients were booted from a server using the local hard disk and APA. APA was able to reduce the boot time of 100 thin clients from 49 minutes to just 12.4 minutes, nearly four times faster.

The Bigger Truth

Improving application performance has long been a key objective of IT. In many of today's environments—including virtualized data centers, private clouds, and public clouds with multi-tenant hosting—the consolidation of workloads both increases I/O traffic and randomizes I/O. High-powered, multi-core server processors and 10 Gb Ethernet ensure that server and network resources are plentiful, but this ability to drive more data through the system leads to storage I/O bottlenecks. Scaling a virtual server deployment is great for the server budget, but with high-density virtualized infrastructures, performance is difficult to sustain within the storage budget.

To handle the performance needs of rapidly proliferating virtualized applications, VDI boot storms, and the demands of cloud computing, many organizations purchase Fibre Channel and/or high-speed SAS HDDs with SSD as a tier of cache or deploy all flash arrays. These are expensive storage options to address a performance problem that can be solved right at the server. PrimaryIO APA enables organizations to non-intrusively accelerate application performance regardless of the shared storage platform. This can mean not only dramatic savings in storage capital costs, but also significantly lower power, cooling, and floor space costs that come from using fewer, larger capacity storage arrays. Back-end storage arrays are freed from servicing high performance read/write operations and can focus instead on storage-based tasks such as snapshots, remote mirroring, compression, deduplication, and backup.

ESG Lab has confirmed that PrimaryIO APA is a hardware-agnostic, software-only accelerator that leverages application-specific algorithms and server-side cache to provide cost-efficient, sustained performance acceleration in cross-platform physical and virtual environments. APA was able to accelerate applications by targeting high-value I/O in a way that is not possible with traditional block-level or LUN level caching. APA also proved that it could enhance the longevity of flash drives even as it boosts performance. APA optimizes SSD wear-leveling and garbage collection and minimizes the number of writes to flash drives accordingly.

PrimaryIO APA offers a cost-efficient solution that plugs into existing infrastructures to provide immediate relief for public and enterprise cloud deployments challenged by growing virtualization. Application-level intelligence enables APA to effectively identify and service the most important data in high-performance mode from a low-latency cache residing on the host, while shared storage can remain configured to maximize capacity and minimize cost.

ESG Lab believes that with APA, PrimaryIO has provided an efficient solution for application performance acceleration that can significantly reduce storage costs for businesses of all sizes.

Learn More about PrimaryIO APA

PrimaryIO.com