OCZ STORAGE SOLUTIONS | A Toshiba Group Company

White Paper

Driving Big Data with OCZ Enterprise SSDs

Part 2: Delivering the Performance and Management Required for Big Data Applications

Scott Harlin

Contents

1 Introduction

In Part 1 of this white paper entitled, "Supporting Big Data Applications with Flash-Based Storage" we introduced key concepts and characteristics associated with Big Data applications to provide a better understanding of this enterprise storage opportunity. We also addressed how flash-based solid-state storage fits into the Big Data model.

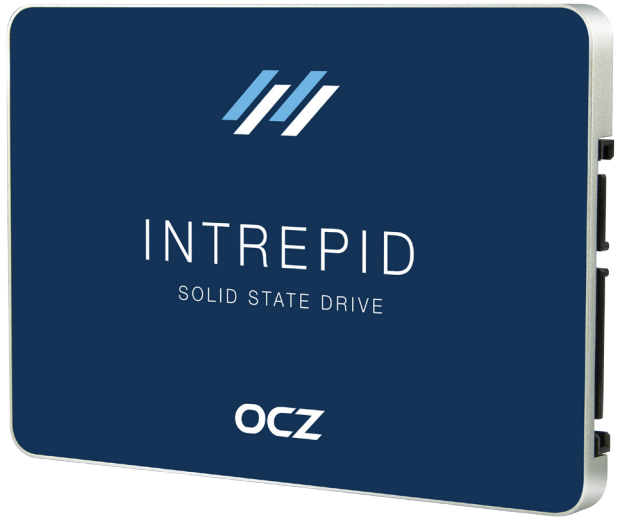
Flash-based SSDs have become the popular choice for Big Data applications as they provide faster I/O performance than HDD storage, support large storage capacities and a variety of form factors and interfaces, consume less power and retain data when power is removed. To gain value from Big Data and achieve a significant return on investment (ROI), IT departments must choose alternate ways to process and analyze data since conventionally the data is too large, moves too quickly or doesn't fit the database architecture structures.

Big Data applications use mixed read and write workloads that require very low latency and significant input/output operations per second (IOPS) performance which is not a good match for hard disk drive (HDD) storage but is ideally suited for enterprise-class solid-state drives (SSDs). In Part 2 of this white paper, we include an overview of OCZ enterprise SSD and software solutions that best address Big Data applications and the ability to deliver ultra-fast processing of large datasets that enable data-driven analytics. The OCZ solutions covered in Part 2 of this white paper include:

- Intrepid 3000 SATA SSD Series
- Z-Drive 4500 PCIe SSD Series
- Windows Acceleration (WXL) Software
- VXL Virtualization Software
- ZD-XL SQL Accelerator
- StoragePeak 1000 Central Management

A quick synopsis regarding how each solution addresses Big Data applications now follows:

2 Intrepid 3000 SATA SSD Series



Designed in a standard 2.5" format in two configurations (Intrepid 3600 cMLC and Intrepid 3800 eMLC) supporting 100GB, 200GB, 400GB and 800GB usable capacities and 19nm MLC flash process geometry

As an HDD replacement, OCZ's Intrepid 3000 SSD Series are ideally suited for Big Data applications representing the Company's highest performing and largest capacity enterprise SATA SSDs to date. The series supports current 19 nanometer (nm) NAND flash process geometries and storage capacities up to 800GB and based on OCZ's Everest 2 platform featuring advanced flash management and endurance capabilities that extend NAND flash life and enhance drive endurance.

Since large mixed read and write workloads, low latency and significant IOPS performance are the basis for Big Data applications, the Intrepid 3000 Series is available in two distinct configurations that address cost-efficient read-centric applications (Intrepid 3600) as well as write-intensive or mixed workload applications (Intrepid 3800):

- **Intrepid 3600**: features reliable and cost-effective Multi Level Cell (MLC) NAND media designed for read-intensive applications such as online archiving, media streaming and web browsing

- **Intrepid 3800**: features high endurance enterprise MLC (eMLC) NAND media designed for write-intensive or mixed workload applications such as Big Data, cloud computing, OnLine Transaction Processing (OLTP), Virtual Desktop Infrastructure (VDI), email servers and analytics

Intrepid 3600/3800 models are based on 100GB, 200GB, 400GB and 800GB usable storage capacities, in 2.5-inch industry standard form factors. In a

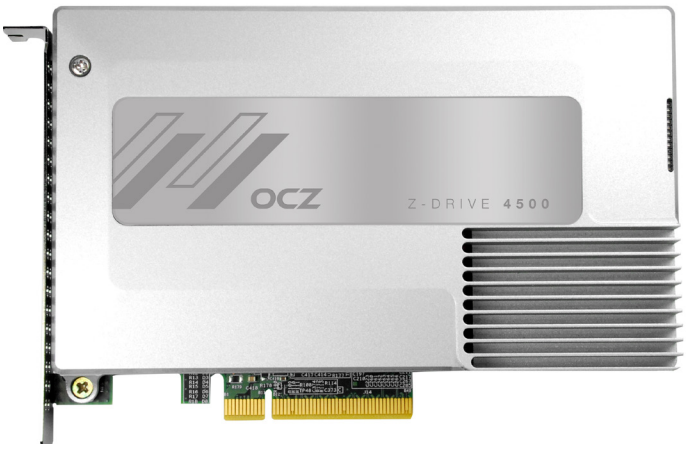
steady state condition by which an Intrepid 3600 or 3800 drive is writing, erasing and re-writing data repeatedly over its full capacity, the performance for both large block sequential operations, as well as small block random operations, is at the top of its competitive class with specifications that include:

520 MB/s for sequential reads (128K blocks)
470 MB/s for sequential writes (128K blocks)
89,000 IOPS for random reads (4K blocks)
40,000 IOPS for random writes (4K blocks)

The Intrepid 3000 Series delivers five times faster sustained performance for 4K write operations and two times faster sustained performance for 4K read operations versus the previous OCZ enterprise-class SATA generation regardless of whether data is in a compressed or uncompressed format. The series also delivers consistent I/O latency so that predictable and efficient I/O performance can be achieved. This level of consistency reduces system and storage bottlenecks, improves end-user productivity, as well as the overall computing experience.

In an identical benchmark test performing a series of small 4K block write operations, the Intrepid 3000 Series consistently improved I/O response times by 12x (or 1200%) over the previous OCZ enterprise-class SATA generation delivering consistent and predictable latency over a sustained time period making this product series well-suited for Big Data applications.

3 Z-Drive 4500 PCIe SSD Series



Designed in a Full-Height/Half-Length (FH/HL) format, the Z-Drive 4500 SSD Series supports 800GB, 1.6TB and 3.2TB usable capacities and 19nm MLC flash process geometry

For server-side deployments, OCZ's Z-Drive 4500 PCIe SSD Series is an excellent alternative to SAS/SATA cabling as a single drive fits directly into a server's PCI Express (PCIe) bus. When flash is inside of the host, a Z-Drive 4500 drive becomes a local resource with performance comparable to the IOPS of servers and an excellent storage solution for today's enterprise Big Data environments. This advanced approach not only moves data onto server-side flash to maximize performance and efficiently utilize host resources, but has significant advantages over disk array SAN storage that occupies more rack space and consumes more power.

To gain value from Big Data and achieve a significant ROI, IT departments must choose alternate ways to process and analyze data since conventionally the data is too large, moves too quickly or doesn't fit the database architecture structures.

The Z-Drive 4500 PCIe SSD Series are designed for the demanding performance requirements of today's enterprise Big Data applications. It leverages 19 nanometer MLC NAND flash supporting 800GB, 1.6TB and 3.2TB usable capacities and delivers even higher performance when compared to OCZ's previous enterprise PCIe SSD generation.

OCZ's proprietary Virtualized Controller Architecture™ (VCA) is leveraged within the Z-Drive 4500 architecture which dynamically reorders storage commands and processes them across eight available controllers effectively appearing and acting as one single drive to the host system. By utilizing the full processing bandwidth of eight controllers working in unison, the storage system runs more efficiently while delivering advanced RAID-like performance all within a seamless, easy-to-deploy solution.

With integrated VCA Technology, the Z-Drive 4500 SSD Series delivers leading sustained performance for enterprise-class MLC-based PCIe edge cards based on industry standard small block tests and benchmarks, and include:

2,900 MB/s for sequential reads (4K blocks)
2,200 MB/s for sequential writes (4K blocks)
252,000 IOPS for random reads (4K blocks)
76,000 IOPS for random writes (4K blocks)

As a result, Z-Drive 4500 models are ideally suited for I/O read and write intensive enterprise Big Data applications where high storage capacities coupled with low power NAND flash results in higher bandwidth and IOPS performance.

4 Accelerating Big Data with OCZ WXL Software

OCZ accelerates Big Data applications even further with its Windows Accelerator (WXL) Software -- a flash management and caching solution for Microsoft Windows Server applications that enables IT managers to deliver low-latency flash deployable as a local flash volume, a flash cache for HDD volumes or as a combination of both. Each model within the Intrepid 3000 SATA SSD Series and the Z-Drive 4500 PCIe SSD Series are supported by WXL Software.

For those Windows applications with small file sizes, data can be efficiently stored on Intrepid 3000 or Z-Drive 4500 flash volumes to take advantage of high-speed flash memory performance. For larger Windows data files that do not fit entirely in flash volumes, OCZ's proprietary cache decision and analysis

By caching hot data on an Intrepid 3000 or Z-Drive 4500 drive, access times are reduced, the Big Data application spends less time waiting for data, and SAN resources are not tied up, increasing Big Data application performance while reducing storage costs and latency-related bottlenecks.

technology makes intelligent selections of what data to store in flash cache. WXL Software is designed to dramatically improve the performance and latency of SAN and DAS systems by intelligently caching the most frequently accessed data on flash storage. By caching hot data on an Intrepid 3000 or Z-Drive 4500 drive, access times are reduced, the Big Data application spends less time waiting for data, and SAN resources are not tied up, increasing Big Data application performance while reducing storage costs and latency-related bottlenecks.

When deployed for caching, WXL Software performs statistical 'out-of-band' processing of all data requests to and from the SAN or internal HDDs using application-specific caching policies that reduce external traffic by up to 90%, storing critical data locally on either an Intrepid 3000 SSD or Z-Drive 4500 PCIe edge card. The caching policies use advanced cache algorithms that detect data hot zones and the most frequently accessed data to be cached while filtering out cold zones so that SSD caching efficiency and endurance can be maximized. WXL Software dynamically distributes flash resources enabling the cache to be shared with other applications so it is accessible by any accelerated volume on the host.

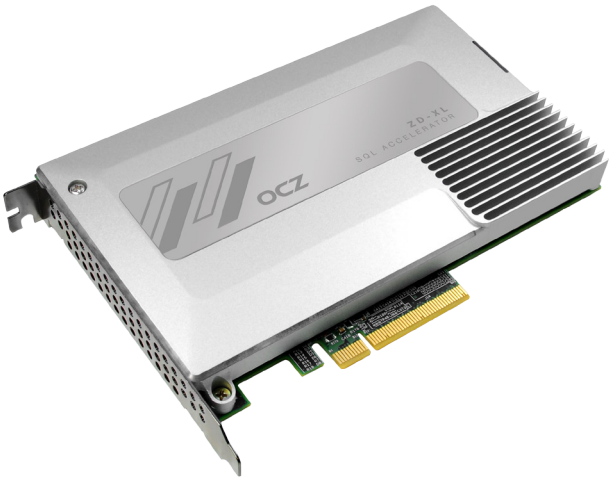
A cache warm-up and analysis mechanism is also featured enabling important and demanding Big Data analytical jobs to be loaded onto the Intrepid 3000's or Z-Drive 4500's flash cache in advance to assure that the critical data is available to the application at the exact time the application needs it.

5 Virtualizing Big Data with OCZ VXL Software

OCZ's VCA Technology is virtualization access technology, and when added to a Z-Drive 4500 PCIe SSD Series, the built-in controllers can efficiently distribute the random loads between all available NAND flash cells to increase and maximize application performance. To take this even further, when a Z-Drive 4500 SSD is combined with OCZ's VXL Virtualization Software, a complete virtual performance system is enabled that efficiently distributes the 4500 flash resources across virtual machines (VMs) to maximize performance of key applications, such as Big Data.

VXL Software enables Z-Drive 4500 PCIe cards to be virtualized as a highly available network resource that allows the flash to be exposed to any VM in a virtualized cluster without negating any of the virtualization services of the hypervisor layer (such as end-to-end mirroring, High Availability (HA), Fault Tolerance (FT) and dynamic VM migration). This advanced virtualization software distributes the flash between VMs based on need making sure that no VM inefficiently occupies flash when it could be better used elsewhere in the environment.

This virtualized approach to application acceleration provides the highest ROI in a virtualized environment where many VMs share the same flash and often do not reach peak workload requirements concurrently. As a result, the Z-Drive 4500's flash cache is optimally utilized at all times regardless of how many VMs are running concurrently, data traffic to and from the SAN is reduced, and critical data is locally available in the Z-Drive 4500 card for immediate use by VMs delivering successful virtualization of Big Data applications.



Designed in a Full-Height/Half-Length (FH/HL) format, the ZD-XL SQL Accelerator supports 800GB, 1.6TB and 3.2TB usable capacities and 19nm MLC flash process geometry

6 ZD-XL SQL Accelerator

To provide accelerated Microsoft SQL Server performance of large database data sets, OCZ's ZD-XL SQL Accelerator leverages proven PCIe SSD hardware and application-tuned software to deliver low latency flash that can also be deployed as a local flash volume, a flash cache for HDD volumes, or as a combination of both. It provides a potent combination of fast flash performance, a unique cache mechanism that makes advanced and statistically-optimized decisions on what data to cache, a dynamic cache warm-up scheduler that enables workloads to be placed on flash cache in advance of demanding and critical jobs, and a wizard-based GUI that enables DBAs to setup caching policies that optimize performance based on SQL Server workloads.

ZD-XL SQL Accelerator provides optimized and efficient flash acceleration for SQL Server environments through its tight integration of innovative hardware and software elements. It supports SQL Server 2008 R2 and 2012 versions, as well as the new 2014 version released April 1st by Microsoft that builds on the key features delivered in previous SQL Server versions, improving storage performance, availability and manageability. ZD-XL SQL Accelerator enables DBAs to unleash the full power of SQL Server 2014 features, such as flash Buffer Pool Extension (BPE) support, that enables database pages to be accessed faster by loading them directly from flash, and a capability well suited for Big Data applications.

Analysing large data sets has become a key issue in today's enterprises as organizations have gained the ability to get access to Big Data and to make it more useful and meaningful. Advanced tools, such as Panorama Software's Necto Business Intelligence 3.0 product, enable additional data exploration and advanced analysis to be performed in a matter of minutes providing a powerful combination of in-memory performance coupled with advanced data discovery tools.

OCZ's StoragePeak 1000 provides a cross-platform view of a company's enterprise flash resources, connected to network servers, storage arrays or appliances, for centralized management, monitoring, maintenance and reporting.

OCZ is currently partnering with Panorama Software to develop joint solutions that transform business information and database datasets into data-driven insights and business intelligence. The ability to deliver real-time Big Data analytics of data stored in Microsoft SQL Server 2014 databases is an important step in the evolution of Big Data applications and solid state storage that drives it. The joint solution is powered by in-memory engines, self-service interactive analytics, infographics and dynamic dashboards enabling business users to easily access, analyse, visualize, track performance, collaborate with colleagues, and share data for quick, efficient and relevant insights that lead to informed decision-making with minimal IT involvement.

With identical performance specifications as the Z-Drive 4500 series, ZD-XL SQL Accelerator models are ideally suited for I/O read and write intensive SQL Server applications where high storage capacities coupled with low power NAND flash results in higher bandwidth and IOPS performance.

7 Central Management via OCZ StoragePeak 1000

The final piece to the Big Data model enables IT managers to centrally perform mission-critical actions and maximize data center ROI from their enterprise flash resources. This level of remote host and SSD management provides the system information and SSD health that IT professionals need to manage their system and storage resources. Developed as a network-accessible management system, OCZ's StoragePeak 1000 provides a cross-platform view of a company's enterprise flash resources, connected to network servers, storage arrays or appliances, for centralized management, monitoring, maintenance and reporting.

StoragePeak 1000 securely connects to multiple host systems across the network and allows IT managers to centrally monitor and administer their enterprise flash resources from a web-based management interface. Supporting enterprise hosts running Linux and Windows operating systems, and featuring an easy-to-use web-based centralized GUI (graphical user interface), IT managers are afforded specific drive details on performance, reliability and operation. Along with the monitoring functionality, a user configurable alerting systems is provided that enables identification of any potential system and/or storage issues in advance enabling corrective actions to be initiated at an early stage.

The user-friendly StoragePeak 1000 GUI provides:

- A structured group-based view of host and SSD activity throughout the data center
- Critical alert displays and warnings from hosts and connected SSDs
- Simpler and easier SSD installation, management and maintenance
- Fast and easy routine SSD maintenance runs, host system checks and administrative tasks from firmware updates to printing detailed reports

As Big Data represents a large volume of both structured and unstructured data that is too big, moves too fast, or exceeds current processing capabilities, the ability to manage and monitor the data activity and flash resources remotely provides a major benefit to Big Data applications.

8 Summary

OCZ provides a complete portfolio of SSD hardware and storage solutions targeted toward Big Data applications that include:

- Leading enterprise-class SATA and PCIe performance of write-intensive or mixed workload Big Data applications with large storage capacities (the Intrepid 3800 and Z-Drive 4500)
- Leading accelerated PCIe performance of write-intensive or mixed workload SQL Server database datasets with large storage capacities (ZD-XL SQL Accelerator)
- Leading cost-efficient SATA read-centric performance with large storage capacities (Intrepid 3600)
- Leading accelerated performance of Windows applications (WXL Software with Intrepid 3000 SATA SSD Series or Z-Drive 4500 PCIe SSD Series)
- Leading virtualized performance of VMware hypervisors (VXL Software with Z-Drive 4500 PCIe SSD Series)
- Leading centralized SSD management (StoragePeak 1000 with Intrepid 3000 SATA SSD Series or Z-Drive 4500 PCIe SSD Series)

Contact us for more information

OCZ Storage Solutions
6373 San Ignacio Avenue
San Jose, CA 95119 USA

P 408.733.8400
E sales@oczenterprise.com
W ocz.com/enterprise

EMAIL SALES TEAM >

VISIT OCZ ENTERPRISE >

As data continues to grow 40% year-over-year, with 90% of the world's data created in the last two years, one thing has become very clear – every enterprise needs to fully understand Big Data and will soon need to implement storage strategies that address performance, analysis and manageability. When this occurs, OCZ Storage Solutions – a Toshiba Group Company is a vendor to consider for these Big Data storage requirements.