Technical white paper

How to choose the right SSD for your workstation

This paper briefly covers some of the Solid State Drive (SSD) attributes that should be considered when choosing an SSD for your workstation. SSD technology is moving quickly with several industry transitions underway. As these changes are adopted, they will add some confusion to the SSD selection process, yet have the potential to dramatically change the storage and memory hierarchy of our platforms while increasing storage speeds and reducing latencies.

Table of contents

SSD categories

As SSD technologies advance there continues to be an almost unmanageable breadth of devices available to our customers. In order bring some order to the chaos HP Workstations has created the following distinct categories for which we provide solutions:

SSD Category	Form factor / Protocol	Features
SATA	2.5" / SATA	Traditional storage connection and form factor
SATA Enterprise	2.5" / SATA	Higher Endurance, Datacenter features
HP Z Turbo Drive G2	Single M.2 / NVMe	Highest performance (6X SATA)
HP Z Turbo Drive G2 TLC	Single M.2 / NVMe	High performance (4X SATA) at lower cost
HP Z Turbo Drive Quad Pro	Up to four M.2 on Add-in-Card / NVMe	Highest performance, Integrated power-loss-protection
Vendor Unique	Various / Various	SSDs with unique features deemed valuable to the workstation market. Typically delivered to the market identifying the vendor and device

HP Workstations provides SSD solutions in a wide range of categories so that a customer can pick the solution that best meets their needs. The table above is an extremely high-level comparison of categories. HP Workstations recommends comparing specific device characteristics in HP's quick specs when selecting storage for specific workstation platforms. Furthermore, the sections below should be used to help guide customers to the appropriate storage solution.

Performance

The architecture and components associated with storage subsystems can dramatically affect the system performance and a user perceptions related to the interactivity of the system. Choosing the correct storage subsystem is important to delivering a great user experience, yet it is difficult to choose the best solution without deployment and testing within a user's environment. The difficulties stem from the fact that it is hard to ascertain the storage system workload, and how such storage-centric performance metrics as sequential reads, sequential writes, random reads, random writes, IOPs, and queue depth relate to the given workload.

It is always best to deploy a proposed storage solution and test it in real world conditions, yet that may not be practical. In order to determine what to deploy and test, the following guidelines can help.

Performance test methodology

Performance measurements are easy to capture, but hard to replicate with SSDs. The major reason for this is that historical access patterns affect the performance of an SSD. In order to get reliable, consistent performance measures a methodical approach must be taken that includes:

- Adjust the system parameters (BIOS and OS) for high performance
- Purge the device to get back to a known initial state
- Precondition the device with consistent preconditioning steps
- Run the benchmarks repeatedly until a steady state performance level is maintained
- Record the steady state performance level

Several good white papers are available that discuss performance test methodologies, a quick web search will provide the reader with innumerable reading opportunities. Also see the HP Whitepaper: Optimizing performance on Windows 7 for NVMe™ and other storage devices.

Performance specifications

Vendors typically provide information that specifies the peak performance of their SSD. Their performance numbers will specify the maximum bandwidth they can sustain as well as the level of latency that can be expected. These numbers, while providing a good representation of the capabilities for a new "fresh out of the box, (FOB)" SSD, don't necessarily provide a reliable performance metric for a user's workload.

Sequential versus Random

SSD drive vendors provide sequential read/write numbers in Megabytes per Second (MB/S) which represents both the read and write bandwidth of the drive. The size and number of outstanding transactions can vary dependent on the vendor's discretion and is typically representative of the FOB performance level and does not include the overhead associated with a drive that has been in use for a period of time.

Random read/write numbers are provided in Input Output Operations per Second (IOPs). IOPs represent the typical latency of the drive. Again these numbers may have been captured in a FOB condition and may not accurately represent what a user will see when the device is deployed and in use.

Be wary of vendor specifications

It is often difficult to translate drive specifications to user's expectations. The FOB concerns have already been mentioned. Some SSD controllers compress the data files before writing them to the drive. The vendor specifications may be given for the performance of files that can be compressed. There are many file formats that cannot be compressed, and the performance for writing these files will be significantly less. One last consideration is whether the OS and storage configuration supports TRIM. If TRIM is not supported (example: RAID configurations on many SAS controllers), then there will be performance degradation, and the endurance of the drive will be shortened.

Different tests focus on different aspects of the storage subsystem

There are many benchmarks that can give insight into an SSD's performance and capabilities. While the HP Workstation Storage team prefers the trace-based SPECwpc, which attempts to categorize the storage performance based on a workstation market verticals, other benchmarks provide a window into the drives performance and capability. They can help a user further narrow the drive choices. They should not replace the benchmarking of the drive under actual use cases and real workloads. Some SSD benchmarks which can be considered are as follows:

	Benchmark	Compression	Type	Comment
PC client	AS SSD	Incompressible	Synthetic	Tests NAND array performance
	ATTO	Highly compressible	Synthetic	May skew data for those drives that compress
	Anvil	Variable compression	Hybrid	GUI on top of IOmeter
	CrystalDiskMark	Both	Synthetic	Widely used for office workloads
	HD Tune	Both	Synthetic	Used to quantify RAM cache behavior
	IOmeter	Variable compression	I/O workloads and latency	Industry standard used to create workloads. All transactions asynchronous
	PCMark Vantage / 8	Mostly incompressible	Trace-based, office	Most trusted real world
Wkst	SPECwpc	Incompressible	Trace-based, workstation	Workstation market segment individually identified. Uses IOmeter with workloads traced from Workstation applications

Synthetic benchmarks are created to represent a use case or expected workload. They rely on the knowledge of the test writer, or in some instances tester, to understand and create real-life representations of user's workloads.

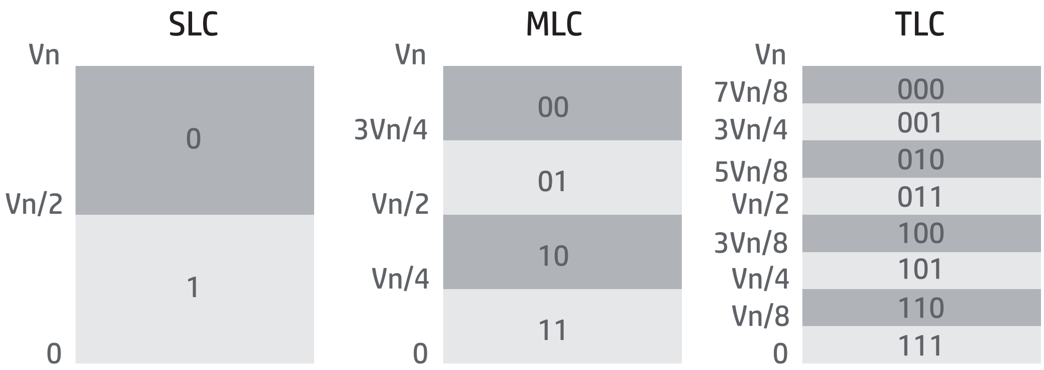
Trace-based benchmarks rely on captured trace information when running specific applications and workloads. They are typically low-level traces of the storage subsystem and may not accurately represent interactions between the storage subsystem and other components in the system.

Hybrid benchmarks are a combination of Synthetic and Trace based benchmarking.

Non-Volatile Media

While there continues to be marketing hype about new media technologies like Intel®/Micron's 3D XPoint the majority of today's SSDs are made with a non-volatile memory technology called NAND. Since NAND makes up the majority of devices on the market this paper will focus on NAND technology. Expect new non-NAND media to have unique characteristics not covered by this white paper.

NAND cells store charge which represents the data state or data value. Each data value has a unique charge state and the more bits per cell the finer the granularity required to set and detect a specific range of stored charge. The figure below shows this concept for Single-Level cells (SLC), Multi-level cells (MLC) and Triple-level cells (TLC). It is important to note that as planar or 2D NAND lithography have decreased the amount of overall charge per cell has decreased. With the advent of 3D NAND, the cell lithography has increased allowing more charge per cell and providing a good path forward for NAND based SSDs.



Endurance

Due to usage, specifically the writing of data, the NAND cells wear, eventually reaching a condition where the cells are unable to store the correct amount of charge. Essentially the cells are worn out. Once enough cells are worn out, the SSD is considered to have reached an end-of-life situation and no more data can be written to the drive. The drive should be replaced prior to this condition.

There is a cyclical nature to how NAND wears out. NAND arrays consist of blocks of data which represent the smallest chunk of the NAND array that can be erased. Each block of NAND consists of multiple pages, each page represents the smallest chunk of NAND that can be written. NAND cells are written one page at a time and can only be written once. They must then be erased at a block level before the pages can be written to again. This process from initially writing a page within a block and then clearing the block so that it can be reused is called the program-erase (P/E) cycle. The P/E rating can vary dramatically depending on the type of NAND being used.

The P/E cycle is the starting point for understanding endurance; it is not the end point as many SSD vendors no longer publish the P/E cycles of the NAND array. The problem is that knowing the P/E rating of NAND does not tell a user how many writes the host can do before the drive is worn out. SSDs are complex devices with firmware that manages the data on the drive and attempts to extend the lifetime of the drive while managing the NAND array to guarantee that it wears evenly. Because of these complexities, vendors specify either a Total Bytes Written (TBW) for client drives or they specify the number of Drive Writes per Day (DWPD) for enterprise drives. Typically TBW = DWPD*365*years of life*capacity of drive. HP recommends enterprise class or datacenter drives for high endurance environments as their DWPD are higher than client drives.

Data Integrity

Data integrity refers to the ability of a drive to assure the accuracy and consistency of the data while it is within the domain of the drive, essentially from the time it reaches the drive to the time it leaves the drive. All drive vendors are concerned with data integrity, yet exact methods for handling it vary dramatically. While there are many aspects to data integrity, this paper covers three fundamental components which should be considered.

End-to-end data protection

End-to-end data protection defines how the data is protected while within the domain of the drive. HP specifically requires that the end-to-end data protection support a minimum of single bit correction and double bit detection on the data path elements within the drive. Many consumer class drivers support detection and not correction which reduces the data integrity of the drive. HP does not require the data at rest to be protected in this manner, although it is expected that all drives support data at rest reliability through techniques like XOR parity striping or other RAID redundancy techniques. Some drives go even further, applying error detection codes at the host interface and storing these codes in the NAND. Storing the codes in the NAND adds an extra layer of verification.

The HP Workstation team picks SSD drives that are designed to protect the data when being read or written (data path), and protect the data when at rest on the drive. The exact algorithms vary by SSD vendor, so it's difficult to provide an extremely precise description that covers all vendors' drives.

Write caching

Write caching is a mechanism to increase the write performance by putting a small and fast DRAM in front of the long-term storage media, NAND in the case of SSDs. The DRAM is used as a temporary storage location until the NAND can be written. This technique works great until power is unexpectedly lost to the drive causing the data in the DRAM to be lost. The data which was presumably committed to the storage subsystem will be lost. Disabling write caching or enabling write cache buffer flushing reduces the concerns about data loss during sudden power loss conditions. Be aware that this can negatively impact storage performance.

Power loss management

Power loss protection (PLP), typically an enterprise feature, is often called out independently from the end-to-end data protection enhances the data integrity of drives that support this feature. Essentially vendors will maintain the power to components within the drive until data is removed from volatile storage locations including the write cache. Typically power is maintained through the use of either capacitors or batteries. Most enterprise drives support PLP, most client drives do not. For client drives used on the HP Z Turbo Drive Quad Pro, HP has developed specialized firmware and an external mechanism for supplying charge on a sudden power loss. This client focused PLP solution is currently only available on the HP Z Turbo Drive Quad Pro.

Reliability

All drives provided by HP workstations have a higher level reliability than off-the-shelf drives. It is expected that our drive vendors provide a higher level of oversight on our drives including: SSD BOM management which controls the quality of the parts used. SSD firmware management, and most importantly more extensive testing including extensive reliability and configuration testing. You can be guaranteed that SSDs sold by HP workstations have been tested in the configurations offered.

Long lifecycle for specific vendor and part number

SSD technology is changing quickly, whether that relates to the number of bits stored in a NAND cell, or the number of NAND planes that can be stacked. While this quick technology transition is in part because vendors are looking at reducing costs, it does have some drawbacks. Users will experience shortened lifecycles for storage components and quicker adoption of new technology, often with a cadence of a year or less.

If maintaining a specific SSD part number is important to you, the lifecycle expectations will need to be understood. If maintaining a specific part number is unimportant, but maintaining a specific level of functionality is important, HP workstations is making every attempt to maintain a level of consistency for categories of SSD devices so that capabilities and features are maintained.

Security

Security as it relates to SSDs typically takes one of several forms:

- None or standard SSDs
- ATA drive lock (SATA SSDs)
- Pyrite (NVMe SSDs)
- TCG Opal 1.0 SED
- TCG Opal 2.0 SED
- IEEE-1667 (eDrive)

Each of these features will be briefly discussed below. If security is required in your environment, the capabilities of your security management software must be considered when choosing an SSD that works with your software. All HP's SSD SED (Self-Encrypting Drive) vendors certify their devices with the major security management software providers, but HP recommends the devices still be tested against a customer's specific security environment before selecting a specific class of device.

All modern SSDs encrypt data but most have no mechanism to support management of an encryption key. Most SATA SSDs do support ATA drive lock which enables password protection of the data on the drive, but not encryption. NVMe SSDs are transitioning to support Pyrite which helps to manage access to logical blocks of the SSD, but similar to ATA drive lock does not encrypt the data on the drive.

TCG Opal 1.0 provides a standardized mechanism to support SSD SEDs on client platforms. TCG Opal 1.0 when combined with security management software provides a more complete security solution. Opal 1.0 features include a pin based access control mechanism for up to four users and a single administrator. It supports encryption over specified LBA ranges so that individuals have uniquely encrypted data spaces. It allows administrators to lock/unlock read and/or write access to specified LBA ranges. It creates an MBR shadow to protect the pre-boot environment.

TCG Opal 2.0 is not backwards compatible to Opal 1.0 and problems have been observed when deploying Opal 2.0 drives in security environments that only support Opal 1.0. Opal 2.0 provides support for up to eight users and four administrators. It has more flexibility in LBA range alignment. It also supports an expanded set of storage management use cases and manageability options. Check with your security management software provider for more details.

eDrive is a Microsoft term used to define a device that supports TCG protocols like Opal 2.0 and IEEE-1667 which defines TCG storage silos and mechanisms for authenticating host attachment of transient storage devices. Microsoft Windows 8 and 10 will recognize eDrives and use BitLocker to manage the drive. If other security management software is used in your environment, it is not recommended that eDrives be deployed.

Manageability

Most manageability features revolve around the security capabilities mentioned in the previous section. Some drive vendors have tools available for their specific drives although there are some restrictions. For example, NVMe drives on Windows 7 do not have access to the Admin queue or status information related to the drive while using the Microsoft NVMe driver. Vendor specific drivers, if available, can help to work around this limitation in the Microsoft driver. Look for the latest drivers on HP's driver support website.

HP Performance Advisor software also has some capabilities, including SSD Wear Level. It is free with workstations, and provides support for all SSDs shipped with HP workstations. If these capabilities are required, the selected tool must be verified to work with the SSD choice. It is expected as SSDs become more prevalent more of these tools will become available.

Conclusion

SSDs are high performance and highly reliable devices, but not all SSDs are the same. Each SSD is optimized around similar but slightly different operating points and features. Some are targeted at low-end PC workflows, and some are targeted at high-end workstation workflows. Customers need to prioritize their needs for Performance, Endurance, Cost, Lifecycle, and Manageability features to accurately determine which SSD is right for their needs. The HP workstation storage team is careful in its selection of SSDs so that customers can choose wisely.

Additional Resources

Below is a list of additional HP White papers which may be beneficial

- SSD Endurance
- Optimizing performance on Windows 7 for NVMe™ and other storage devices
- Choosing storage solutions for demanding imaging applications
- NVMe technology for SSDs

4AA5-1095ENW, August 2017