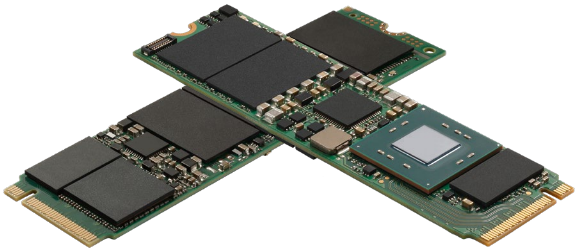
Client vs. Data Center SSDs

A Guide to Understanding Differences in Performance and Use Cases

Overview

Client solid-state drives (SSDs) — those designed primarily for personal computer storage — can excel in some, but not all, data center applications. Data Center SSDs are designed from the ground up for data center use. When considering using a client SSD in an data center application, it is imperative to understand the input/output operations per second (IOPS) performance and design differences between the two.

This brief discusses some of these differences.



Client and Data Center M.2 SDDs

Different SSDs for Different Applications

SSD designers optimize performance and cost based on intended use. Directly comparing SSDs designed for different uses (when examining data sheets, for example) can be difficult. It is like comparing fundamentally different products intended for fundamentally different uses.

We can make more informed decisions when we understand some of the performance implications of using a client SSD in an application for which it was not designed.

Consider an IOPS performance comparison between a client SSD (optimized for personal storage such as mobile computing) and an SSD optimized for mainstream data center use (such as highly active real-time databases).

Because data center SSDs are designed for demanding workloads like this (and client SSDs are not), we expect the data center SSDs to excel (and the client SSDs to falter). A common test illustrating this point is a 4KB random 100% write workload over an extended period. Figure 1 shows how the performance of each SSD type changes with time. FOB is “fresh out of box,” meaning the SSD has experienced few or no program/erase (P/E) cycles yet.

Although the exact shape of these curves may change with different SSDs and workloads, all SSDs undergo this performance change. With this example workload, the data center SSD shows higher steady state performance. Steady state write performance is the most important factor for data center customers.

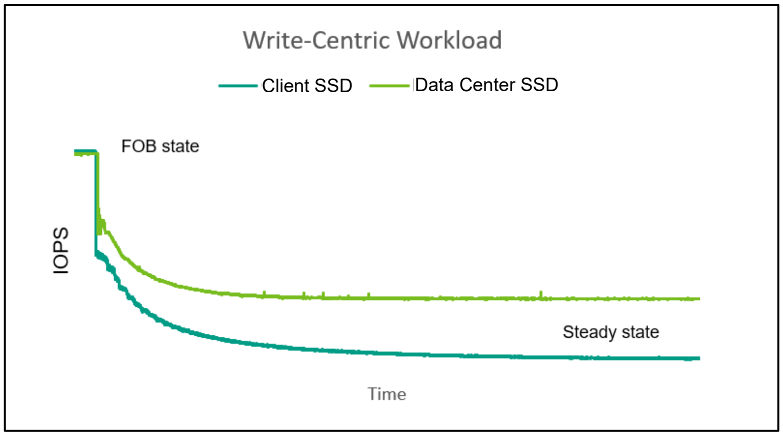


Figure 1: SSD Performance Over Time

It is important to note that the comparison in Figure 1 is only one aspect of drive performance. It is not a complete representation for all applications, uses or standard benchmarks. It illustrates that good performance is relative to the target application and use.

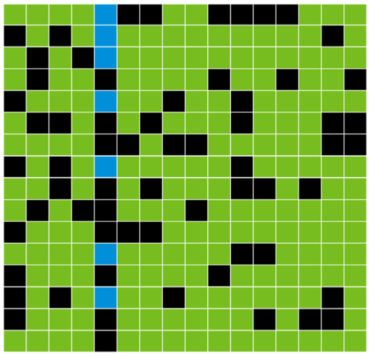


Figure 3: Garbage Collection With 25% Over-Provisioning

Factors Affecting Write Performance: Understanding Over-Provisioning

Over-provisioning is additional media space on an SSD that does not contain user data. Every SSD has some level of over-provisioning.

Figure 1 shows the 4K random write performance of a client and an data center SSD over time. The data center SSD has considerably more over-provisioning. That additional media space plays a critical role in steady state random write performance.

This section explains why.

Introduction to Garbage Collection

When 3D NAND media (the media used in SSDs) has been written, the media must be erased before it can be rewritten. This is different from hard disk drives (HDDs). HDDs use “write in place” media. If the HDD media already contains data, we can overwrite the data in a single step. NAND takes two steps (erase and write).

NAND is organized by pages (the smallest portion that can be written) and blocks (the smallest portion that can be erased). Blocks contain many pages (the exact number depends on the NAND design). When we want to erase a NAND page so we can write new data to it, we cannot erase just that page — we have to erase an entire block. If the block has some data we want to keep, we have to move that data by writing it somewhere else on our SSD before we erase the block.

A process known as garbage collection accomplishes this in two steps. The first step identifies the data we want to keep and moves it to a free location on the SSD. Once complete, the second step erases the block to produce pages to which we can write new data.

The example in Figure 2 helps illustrate garbage collection on a hypothetical client SSD. This example contains 256 NAND pages, shown as squares (real SSDs have far more pages), and each column of cells represents a block. The green squares represent pages with data we want to keep. The black squares are pages that are ready to receive new data. The blue cells are pages with data that we need to keep but that we also need to move to erase the block.

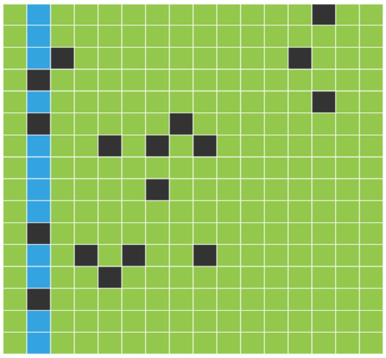


Figure 2: Garbage Collection With 7% Over-Provisioning

This client SSD contains about 7% over-provisioning.

In this example, the SSD must move the data in the blue cells before it erases the block (column). Note that there are not many areas into which the data can be moved (black cells). This is due to limited over-provisioning.

Figure 3 shows a similar example but with an SSD that has 25% over-provisioning. As before, this SSD must first copy the data we want to keep into new pages so it can erase the block.

Note that the amount of over-provisioning affects the amount of data that must be moved for the SSD to erase the block.

In the 25% effective over-provisioning example, the SSD has to move fewer pages before erasing the block. There are also far more areas into which the data can be moved (black cells). This enables better optimization, making garbage collection more efficient.

The 25% over-provisioning example improves SSD write performance for two reasons:

- Fewer pages (with data we want to keep) need to be moved before a block is erased.
- More places to store the data that needs to be moved means better optimization and more efficient garbage collection.

Over-Provisioning and Random Workloads

SSD over-provisioning is calculated as a ratio and expressed as a percentage:

We can see the effect over-provisioning has on IOPS performance when we adjust over-provisioning on the same data center SSD, applying the same random workload iteratively.

Figure 4 shows how different over-provisioning levels can affect IOPS performance. In the example, we performed the same test on the same data center SSD containing the same firmware installed in the same system. We only varied the level of over-provisioning (OP).

For these tests:

$$\frac{\text{Total media space}}{\text{Media space available for data storage}}$$

- We restored the SSD to FOB before we started each test and applied a small transfer, random, mixed IO workload.
- We started with the default capacity (blue) and then increased the over-provisioning using Micron's Flex Capacity feature to +17% (over default) and then +50% (over default).

Figure 4 shows the test results:

- Additional over-provisioning increases the IOPS performance at steady state.
- It does not affect IOPS performance at FOB.

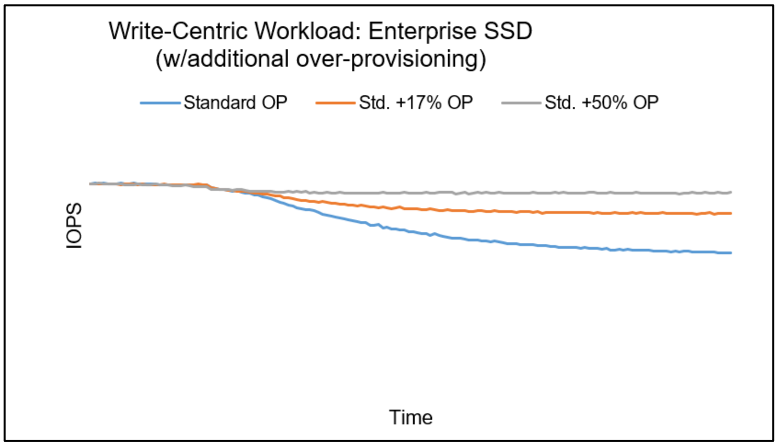


Figure 4: Effects of Adding Over Provisioning

The numbers may change based on the drive and workload tested. The relative results and overall principle remain the same: Increasing over-provisioning (even on a data center SSD) improves IOPS performance for workloads with a write component (mixed I/O).

Here is why: As the write amplification decreases, the random steady state performance improves.This is because of the improvements in garbage collection efficiency, as discussed in the previous section.

Over-Provisioning and Sequential Workloads

Sequential workload IOPS performance is affected far less by changing over-provisioning levels than random workloads are. This is because sequential workloads place the data in a more orderly manner as they write it. Figure 5 illustrates this process. Using the same hypothetical example SSD, Figure 5 shows an example of data placed by a sequential workload. Because the data is more orderly (compared to random workload placement), garbage collection does not happen as frequently.

Client and data center SSDs typically show good sequential workload performance.

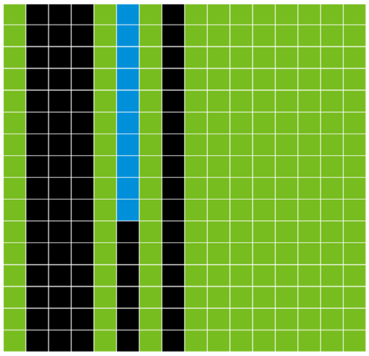


Figure 5: Garbage Collection on a Sequential Workload

Write Buffering and Steady State Performance

Traditionally, write buffering has been used to increase instantaneous, or burst, I/O performance. Incoming write traffic is buffered into very fast storage (usually DRAM) and then migrated to slower, long-term storage (NAND). Because buffers are typically limited in size, they have not been regarded as a factor in steady state performance. Once the buffer fills, it brings no benefit. (To absorb an incoming write, we must drain data from the buffer into the NAND.)

For client and data center SSDs, the write buffer may improve steady state IOPS performance. This is because SSDs extensively use parallelism to improve IOPS performance. If we can increase parallelism, we increase IOPS performance.

One method for increasing parallelism is write accumulation. Write accumulation is a process by which several smaller write operations are combined into a larger write operation across multiple physical NAND die.

This process optimizes write operations: It enables the greatest amount of data to be written with the least amount of media busy time.

To take advantage of write accumulation, the SSD must have some form of write buffer in which to accumulate write commands.

Although client and data center SSDs can use this technique, the exact implementation may differ. Micron data center SSDs have stored energy to write all the data in a write accumulation buffer to NAND when the SSD loses power (due to sudden removal, for example). Without a power protection mechanism, this sudden power-loss may result in data risk.

Typical client SSDs do not have this capability. This is because, in conventional personal storage applications such as personal computing, this difference is inconsequential. (The SSD cannot be removed without powering the system down. If it is, the operating system also halts because it, too, is stored on the SSD.) One can disable the write buffer on client SSDs, but performance may be reduced.

Power-Loss Protection

Client and data center SSDs both use nonvolatile NAND memory for long-term data storage. Different types of NAND store a different number of bits in each cell. Single-level cell NAND stores one bit per cell; multilevel cell

(MLC), two bits per cell; triple-level cell (TLC), three bits per cell; and more recently, quad-level cell (QLC), four bits per cell. The more bits per cell, the higher the NAND (and drive) density.

MLC, TLC and QLC NAND have some limitations. For example, these devices can be vulnerable to data loss in the event of an unexpected power loss. This white paper on micron.com discusses this phenomenon in detail.

Client and data center SSDs have different levels of power-loss protection (PLP). Client SSDs protect data at rest. Data center SSDs protect data at rest and data in flight. “Data at rest” is data that has been successfully written to the storage media. “Data in flight” refers to data that has been sent to and acknowledged by the SSD (but may not yet be committed to the media, such as data temporarily buffered in volatile memory) or any write that is in progress but not yet complete.

Client SSD Power-Loss Protection — Data at Rest

For typical client SSD use, data at rest protection is usually sufficient. Figure 6 shows client SSD PLP, extending to data already stored in the nonvolatile media (gray). Figure 7 shows data center PLP, which extends from the nonvolatile memory (as in client PLP) through the volatile memory to protect committed writes not yet stored in nonvolatile memory, as well as writes to nonvolatile memory already in process.

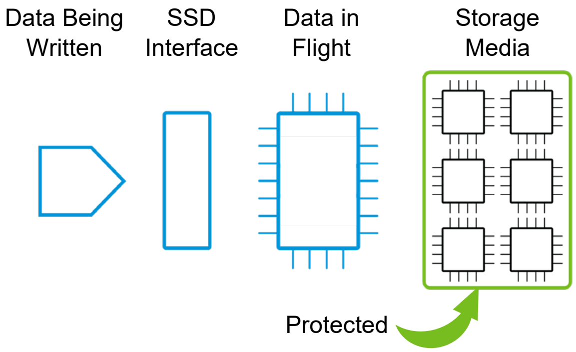


Figure 6: Client Power-Loss Protection

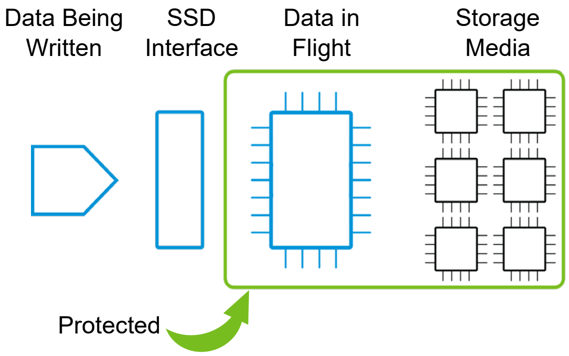


Figure 7: Data Center Power-Loss Protection

Data Center SSDs have extended PLP because data loss in the data center is more critical than in client computing. Client devices are typically single user, so while data loss is important, it affects only one user. In addition, the amount of unprotected data in flight is typically small, less than 2MB. Modern desktop applications are often able to compensate for this small risk by journaling the user’s activity so that unsaved changes can be recovered in the event of an unexpected power loss.

Data Center SSD Power-Loss Protection — Data at Rest and Data in Flight

On the other hand, data center SSDs are often installed in platforms supporting hundreds of users and mission-critical systems. Data loss here potentially affects hundreds of users or more and can have greater consequence. With data center SSDs, is it essential to protect data at rest, like in client SSDs, but also data in flight. Any writes in progress must be completed, and any data buffered in volatile memory must be committed to the NAND device and protected.

Summary

Many factors affect SSD performance in a given application. How the application accesses the SSD (randomly or sequentially) can influence SSD IOPS performance, as can the basic design of the SSD itself.

It is important for system designers to understand some of the key differences between client and data center SSDs to ensure an optimal fit for their use models.

micron.com

Rev. C 07/2021 CCM004-676576390-10823