SAMSUNG

White Paper:

# SSD Technology: How Over-Provisioning Impacts Performance

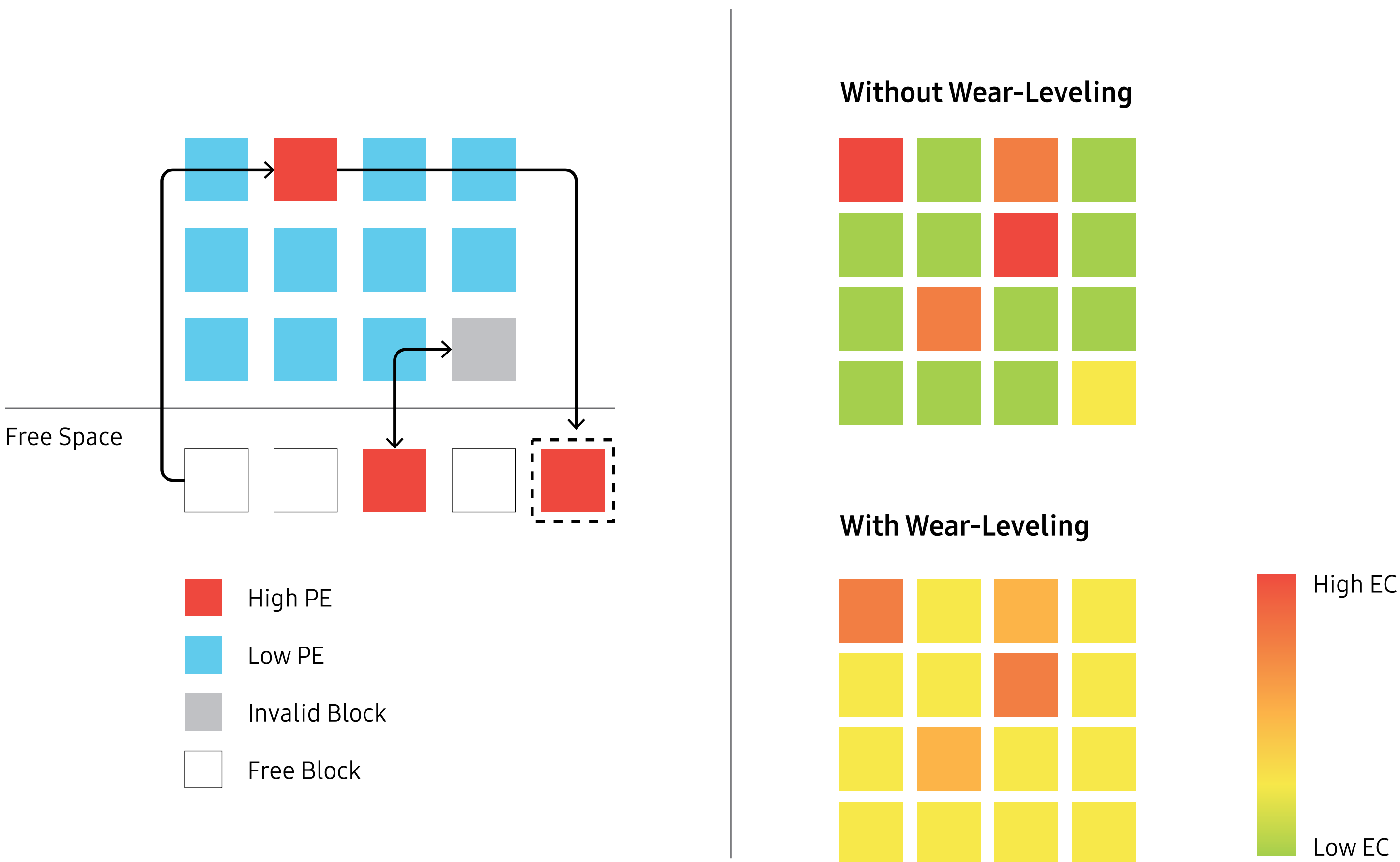
# Introduction

Unlike hard disk drives (HDDs), solid state drives (SSDs) store electrons on NAND cells for writing data, which cannot be overwritten when the data is being stored or erased. Since the storing and deleting operations (program/erase) of an SSD are carried out in different units — pages and blocks, respectively — a repeated sequence of program and erase is inevitable in writing and managing appropriate data.

As more of these program/erase (P/E) cycles repeat, some electrons become trapped between cells, in which case, if it worsens, the corresponding cells reach the end of their lifetime and decrease overall durability of the storage device. This phenomenon is called the wear-out of cells and is responsible for the physical limits of the lifetime of NAND.

Therefore, good management of the NAND lifetime is crucial in elongating the usage of the SSD. In other words, since the corresponding cells quickly wear out when data is repeatedly written in a certain region, the reuse of certain cells should be prevented. This function, called wear-leveling, provides an even usage of all cells by swapping blocks of high P/E cycles with blocks of free space, allowing the user to keep the SSD for a long time under given conditions. (Figure 1)

**Figure 1:** Wear-Leveling



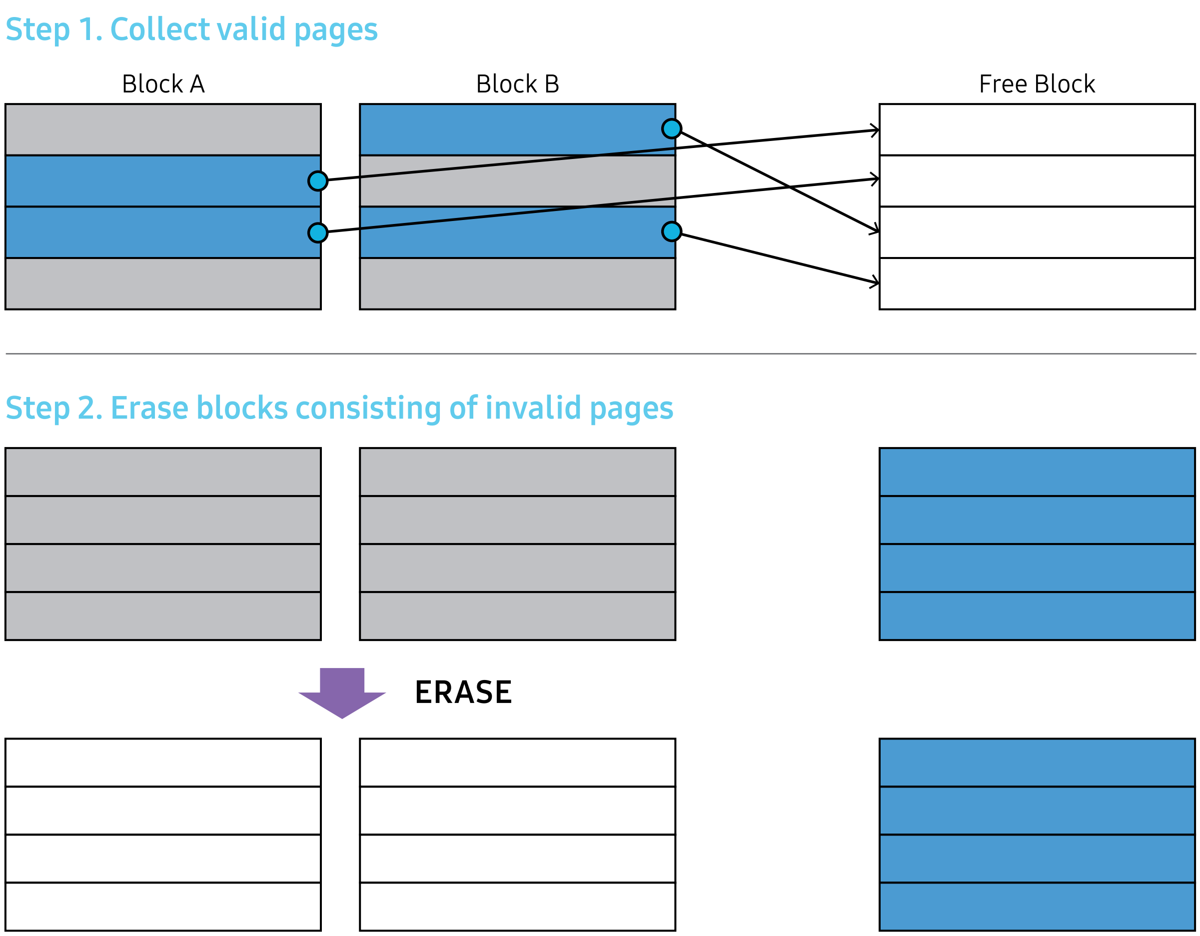
# Garbage Collection

The erase operation that results from data's inability to carry out the overwrite function can also influence the writing performance. This occurs most commonly due to the longer time required for the erase operation compared to the write operation itself. As mentioned above, the write function is carried out by pages while the erase function operates in blocks.

To improve this issue in performance, a process called garbage collection (GC) is applied to create free blocks within the SSD. This technology secures free blocks by collecting valid pages into a single location and erasing the blocks consisting of invalid pages. However, a slower performance is possible in the case of an unexpected situation where the GC interferes with the host write. The SSD therefore requires free space to allow the firmware feature to run smoothly. The creation of these free blocks is called over-provisioning. (Figure 2)

**Figure 2:** Garbage Collection

Valid Page | Invalid Page | Free Page



# What Is Over-Provisioning?

Over-provisioning (OP) refers to a function that promotes efficient usage of the SSD by allotting a certain amount of memory for SSD NAND as the OP space, so that it's inaccessible from the host and only the controller has access. Consisting of free blocks only, the OP region assists in efficient delivery of free blocks when wear-leveling or GC is in progress and contributes to improved performance and lifetime of the SSD. Typically, Samsung DC SSD provides 6.7 percent of factory OP as its default, but users can manually adjust the amount of space when they require additional OP depending on the user environment. (Figure 3)

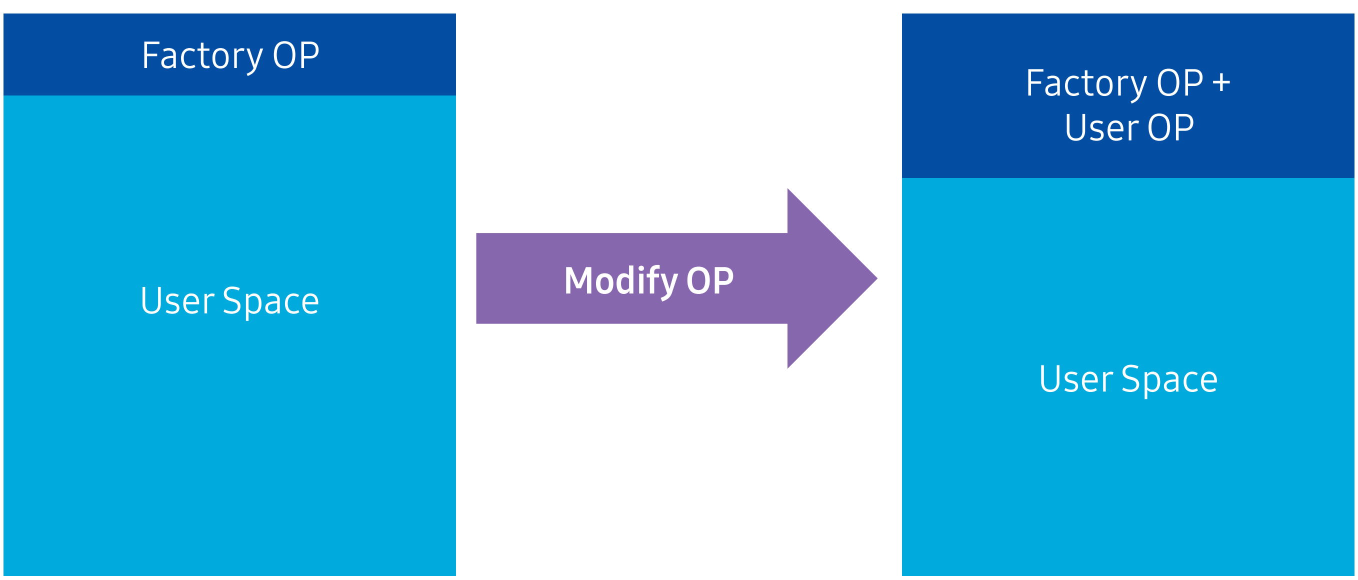
## Calculating the OP Ratio

OP Ratio Formula:

$$\text{OP (\%)} = \frac{\text{(Physical Capacity – User Capacity)}}{\text{User Capacity}}$$

Example: When 120GB of a 128GB SSD is used as the User space while 8GB is assigned to the OP, the OP (%) is ((128 – 120) / 120) x 100 = 6.7%


**Figure 3:** Over-Provisioning



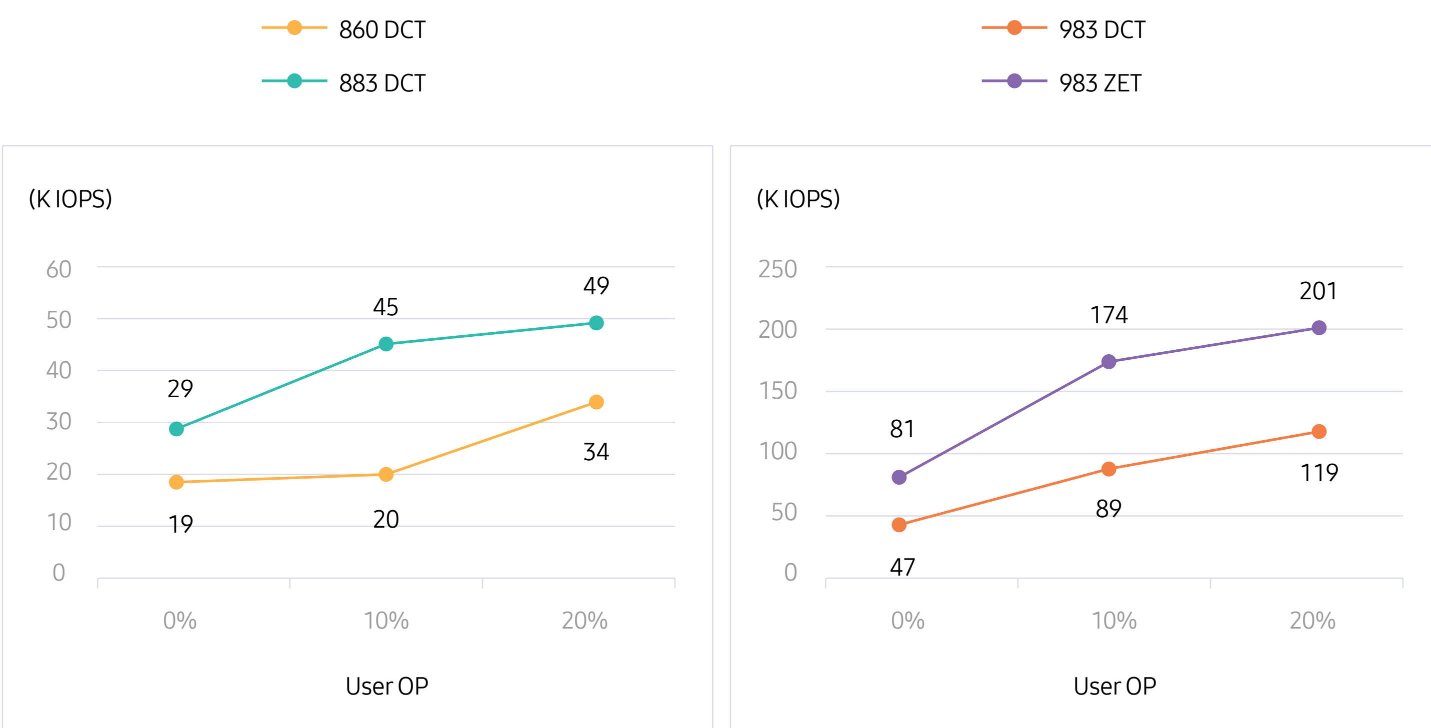
# Advantages of Increasing OP

Although there is no difference in the performance of sequential and random writes for fresh-out-of-the-box (FOB) SSDs, once data has been written over the entire space of NAND, the random write does not perform as well as the sequential write. The random write, smaller than the sequential write, mixes the valid and invalid pages within the blocks, which causes frequent GCs and weaker performance. With increased OP, more free space — which the host does not have access to — can be secured, and the resulting efficacy of GC assists in improved performance. The sustain performance is improved in the same manner. (Figure 4)

Internal operations such as GC cause the value of the NAND Write to become greater than that of the Host Write, and this also results in an increase in the Write Amplification Factor (WAF), the ratio of the Host Write to NAND Write. An increase in the WAF value indicates that unexpected NAND usage is increasing, and the product may not last until the total byte written (TBW) is reached.

A sufficient OP space has the advantage of increasing the drive writes per day (DWPD) that users can employ during the warranty period by decreasing the NAND usage through the improvement in the efficiency of internal workload of NAND.

**Figure 4:** Random Write Performance (QD32) per User OP for SATA / NVMe DC SSD



# Samsung's DCT Series

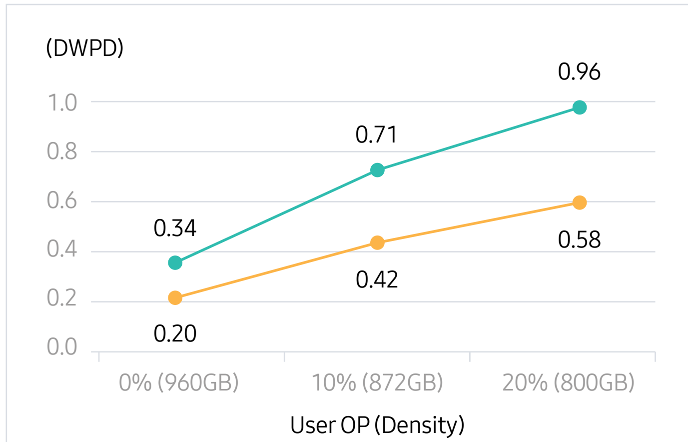
The graphs below show the evaluations of DWPD per warranty by allotting the additional User OP for Samsung's Data Center products (860/883/983 DCT, 983 ZET). Each value in the graphs is calculated through the formula below and is a measure value of each SSD, not a guaranteed value. The value of DWPD grows in accordance with the increase in OP rate. The user can check the numbers that are needed for the calculations below through smart attributes. (Figure 5)

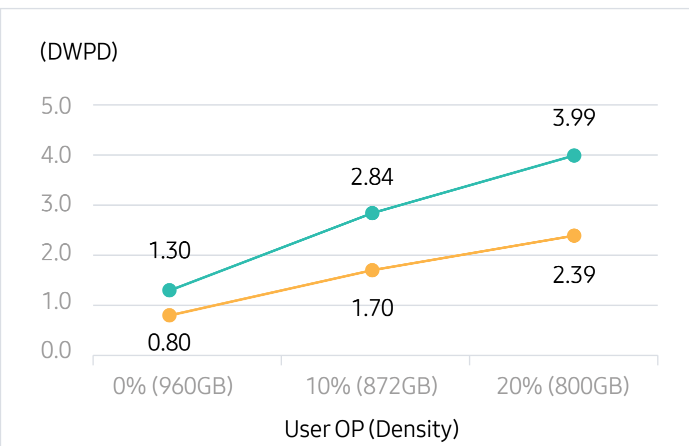
$$\text{WAF} = \frac{\text{Physical Write Amount}}{\text{Host Write Amount}}$$

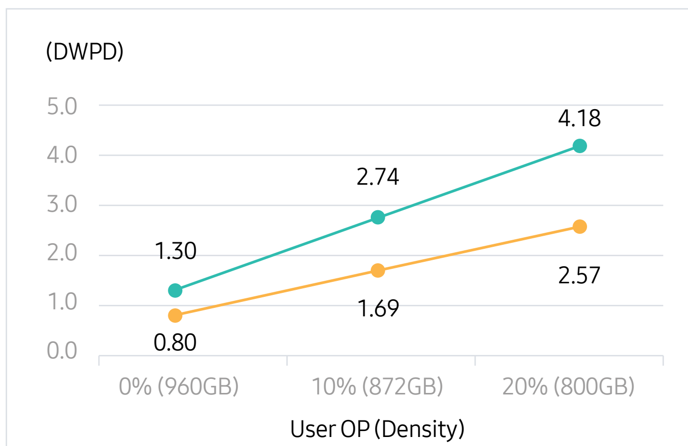
$$\text{DWPD} = \frac{\text{NAND PE Cycle + Raw Density}}{\text{Logical Density x 365 x Warranty Years x Waf}}$$

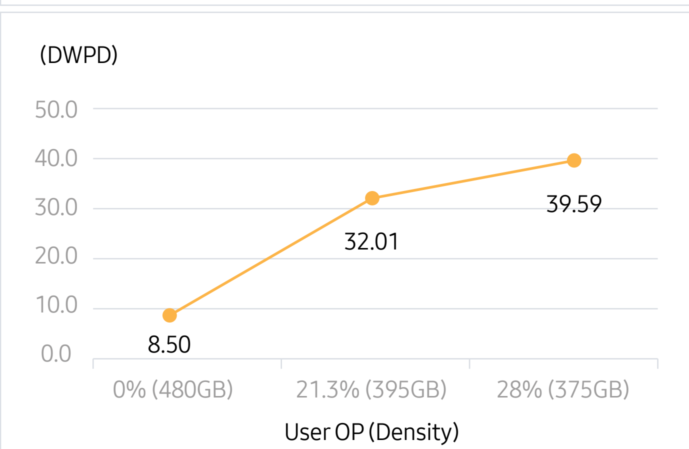
**Figure 5:**
The Calculations

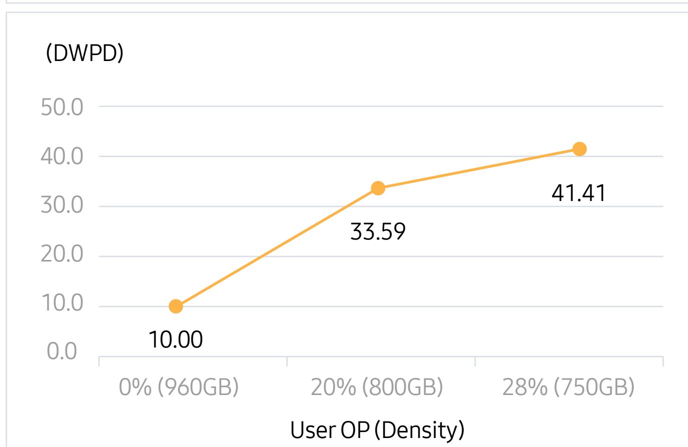
Warranty 3yrs
Warranty 5yrs



# Estimating Lifetime of an SSD Using Smart Attributes

OP refers to a function that promotes efficient usage of the SSD by allotting a certain amount of memory for SSD NAND as the OP space, so that it's inaccessible from the host and only the controller has access. Consisting of free blocks only, the OP region assists in efficient delivery of free blocks when wear-leveling or GC is in progress and contributes to improved performance and lifetime of the SSD. Typically, Samsung DC SSD provides 6.7 percent of factory OP as its default, but users can manually adjust the amount of space when they require additional OP depending on the user environment.

| ID | Attribute name | Status Flag | Threshold (%) |
|---|---|---|---|
| **177** | Wear-Leveling Count | 010011 | 5 |
| **241** | Total LBAs Written | 110010 | - |
| **242** | Total LBAs Read | 110010 | - |
| **245** | Timed Workload Media Wear | 110010 | - |
| **246** | Timed Workload Host Read / Write Ratio | 110010 | - |
| **247** | Timed Workload Timer | 110010 | - |

## Example

A user has witnessed that the usage pattern of his SSD has recently changed from 80 percent to 70 percent read I/O operations and he would like to understand what impact this change has on the lifetime of his SSD. He has decided to run a test for one week. At the end of his test run the smart attributes read as follows:

ID 245: 4
ID 246: 70
ID 247: 604,800
(7 days x 24 hours x 60 minutes x 60 seconds)

To find the estimated end of life given the above readings, the user would need to do the following calculations:

First, the user would want to understand how many more cycles could be run under the given test scenario before the SSD would wear out completely. He would therefore calculate **1,000 / 4 = 250**

The user would then multiply this number by the duration of the test run to find the total expected lifetime of the SSD in seconds. This calculation would yield **250 x 604,800 = 151,200,000**

Given the relatively abstract nature of large numbers expressed in seconds, the user would then want to express the lifetime in years, months or weeks. If we chose to express the lifetime in years, we would make the following calculation:
**151,200,000 / (365 x 24 x 60 x 60) = 4.79 years**

- **ID 241 and ID 242** indicate the write amount of the Host and NAND, respectively, and users can calculate the WAF of their SSD with these values.
- **ID 177** refers to the number of wear-leveling operations and can also be interpreted as the overall average P/E cycle. With this and the WAF value, users can determine the DWPD.
- **ID 247** represents the seconds that the SSD has been in operation since the workload timer was started. (Users can start/stop said timer at their discretion or let it run continuously. It is controlled through their SSD software tools.)
- **ID 246** shows the share of input/output (I/O) operations that are read commands since the workload timer (ID 247) was started and is measured as a percentage. (Conversely, the share of write I/O operations can be determined by subtracting the given smart attribute reading from 100.)
- **ID 245** measures the wear of the SSD given the workload (ID 246) and the period of time over which these workloads have been sustained (ID 247). It is displayed as a "per mile" reading of the total wear of the SSD over its useful lifetime (e.g., a reading of 1,000 would mean that the SSD has been worn out over the given time and usage pattern).

# How to Adjust OP

Users can use either DC Toolkit or Linux HDParm Disk Management to adjust the available space for the OP in the SSD user space from the OP preferences, but the space that is already in use cannot be adjusted. If the user wishes to increase the OP, then they must clear up some space that is already in use to increase available space for the OP.

| Step | Description | CMD |
|---|---|---|
| 1 | Identify the device connected to the system | DCToolkit.exe -L |
| 2 | Check the MAX ADDRESS setting available range | DCToolkit.exe -d 1 -M -r |
| 3 | Set MAX ADDRESS value | DCToolkit.exe -d 1 -M -s 12345678 |
| 4 | Confirm set value | DCToolkit.exe -L |

Select Administrator: Command Prompt

```
C:\Users\hyo\Desktop\DCToolkit_V2.1.W.9.0>DCToolkit.exe -L
================================================================================
Samsung DC Toolkit Version 2.1.W.9.0
Copyright (C) 2017 SAMSUNG Electronics Co. Ltd. All rights reserved.
================================================================================
--------------------------------------------------------------------------------------------------------------------------------
| Disk   | Path               | Model                     | Serial         | Firmware | Optionrom | Capacity | Drive  | Total Bytes | NVMe Driver |
| Number |                    |                           | Number         |          | Version   |          | Health | Written     |             |
--------------------------------------------------------------------------------------------------------------------------------
| *0     | \\.\PHYSICALDRIVE0 | SAMSUNG SSD 863a          | S361NX0H500008 | GXT51M3Q | N/A       |   894 GB | GOOD   | 0.98 TB     | N/A         |
--------------------------------------------------------------------------------------------------------------------------------
| 1      | \\.\PHYSICALDRIVE1 | SAMSUNG MZ7LH3T8HMLT-00003 | ABCDEFGHIJKLMN | HXT70F3Q | N/A       |   447 GB | GOOD   | 0.00 TB     | N/A         |
--------------------------------------------------------------------------------------------------------------------------------

C:\Users\hyo\Desktop\DCToolkit_V2.1.W.9.0>DCToolkit.exe -d 1 -M -r
================================================================================
Samsung DC Toolkit Version 2.1.W.9.0
Copyright (C) 2017 SAMSUNG Electronics Co. Ltd. All rights reserved.
================================================================================
--------------------------------------------------------------------------------
Disk Number: 1 | Model Name: SAMSUNG MZ7LH3T8HMLT-00003 | Firmware Version: HXT70F3Q
--------------------------------------------------------------------------------
Native SET MAX value of the disk is 937703087 LBAs.
--------------------------------------------------------------------------------

C:\Users\hyo\Desktop\DCToolkit_V2.1.W.9.0>DCToolkit.exe -d 1 -M -s 12345678
================================================================================
Samsung DC Toolkit Version 2.1.W.9.0
Copyright (C) 2017 SAMSUNG Electronics Co. Ltd. All rights reserved.
================================================================================
--------------------------------------------------------------------------------
Disk Number: 1 | Model Name: SAMSUNG MZ7LH3T8HMLT-00003 | Firmware Version: HXT70F3Q
--------------------------------------------------------------------------------
Disk Capacity updated to 5GB.
SET MAX Operation Completed. PowerCycle the disk.
--------------------------------------------------------------------------------

C:\Users\hyo\Desktop\DCToolkit_V2.1.W.9.0>DCToolkit.exe -L
================================================================================
Samsung DC Toolkit Version 2.1.W.9.0
Copyright (C) 2017 SAMSUNG Electronics Co. Ltd. All rights reserved.
================================================================================
--------------------------------------------------------------------------------------------------------------------------------
| Disk   | Path               | Model                     | Serial         | Firmware | Optionrom | Capacity | Drive  | Total Bytes | NVMe Driver |
| Number |                    |                           | Number         |          | Version   |          | Health | Written     |             |
--------------------------------------------------------------------------------------------------------------------------------
| *0     | \\.\PHYSICALDRIVE0 | SAMSUNG SSD 863a          | S361NX0H500008 | GXT51M3Q | N/A       |   894 GB | GOOD   | 0.98 TB     | N/A         |
--------------------------------------------------------------------------------------------------------------------------------
| 1      | \\.\PHYSICALDRIVE1 | SAMSUNG MZ7LH3T8HMLT-00003 | ABCDEFGHIJKLMN | HXT70F3Q | N/A       |     5 GB | GOOD   | 0.00 TB     | N/A         |
--------------------------------------------------------------------------------------------------------------------------------
```

# The Samsung DCT Lineup


## 860 DCT

2.5-in. form factor, capacities ranging from 960GB to 3,840GB, with 550MB/s sequential read speed and 520MB/s sequential write speed

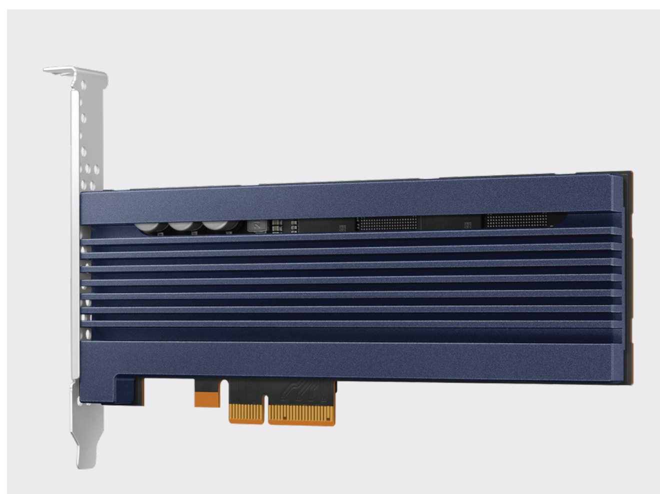

## 883 DCT

2.5-in. form factor, capacities ranging from 240GB to 3,840GB, with 560MB/s sequential read speed, and 520MB/s sequential write speed


## 983 DCT

M.2 form factor, capacities ranging from 960GB to 1,920GB, with 3,000MB/s sequential read speed and 1,430MB/s sequential write speed

2.5-in. form factor, capacities ranging from 960GB to 1,920GB, with 3,400MB/s sequential read speed and 2,200MB/s sequential write speed



## 983 ZET

HHHL form factor, 480GB and 960GB capacities, with 3,400MB/s sequential read speed and 3,000MB/s sequential write speed

# Conclusion

Although increasing OP has the advantage of improving both the performance and lifetime of the user's SSD, it also decreases the available User Space. An appropriate amount of space for the user's workload must be considered when adjusting the OP rather than excessively increasing it. The advantage of the OP yields the best output when the setting of an FOB SSD product is set. Therefore, it is recommended that the user's SSD product is FOB when the OP setting is adjusted.

Learn more about Samsung's entire SSD lineup and how each individual product can impact your business.