# *Enterprise Storage Stack*

## Block Translation Changes the Rules

***ESS*** is a software "block translation layer" from WildFire Storage. It was first deployed in 2005 and it's first patent was issued in 2007.

The idea behind ***ESS*** is to create a storage "device" that has enhanced performance by transforming the layout of writes coming from the application. In general, ***ESS*** creates an environment that presents the storage disks and array with a "perfect write workload".

### What is a Perfect Write Workload?

In general, if you had the ability to design a workload for storage, it would let you define what write pattern and block size would best fit the storage devices and array. While devices and arrays vary, there is a lot of overlap. So what is the perfect write workload?

#### The Perfect Write Workload:

A perfect write workload will write to the device in:

- Long linear writes
  - For SMR HDDs and Zoned Flash SSDs
    Writes that are perfectly aligned with zones
  - For Standard HDDs
    Writes that are as long as practical
  - For Standard SSDs
    Write that are perfectly aligned with the "natural write" size of the SSD.
    This size can take a while to find with testing but is usually around 256 MB for current data center SSDs

A perfect write workload for a RAID set is:

- Writes that are a multiple of the array "optimal IO size"
  This eliminates the read/modify/write in-place penalty of parity-based RAID

### How Can *ESS* Do This?

***ESS*** operates by taking the random write workload from the application and writing it as long linear segments to the media and array. These segments contain not only the data but also control information. The LBA (Logical Block Address) of an application write is no longer tied to a location on the media. System RAM is used to keep a lookup table so that reads access the correct location.

Most SSDs have something similar called an FLT (Flash Translation Layer). ***ESS*** actually pre-dates modern FTLs, and unlike Flash implementations, ***ESS*** can run completely on top of a standard block device.

### *ESS* Impact on Write Performance

***ESS*** writes at the speed of the media. The write pattern of the workload does not really matter, so 4K random writes, and 1M linear writes will both saturate the media and whatever speed the drives can write at. Similarly, a parity-based RAID array will saturate the component drives. Because all IO is at "optimal IO size and alignment", writes involve perfect writes to each member drive with no intermediate reads required.

It is sometimes easiest to consider some examples:

| | |
|---|---|
| 24 QLC SSDs RAID-6: | 350 MB/sec writes for each drive<br>22 x 350 = 7700 MB/sec application-level writes at any block size<br>1.9 Million 4K random write IOPS |
| 24 TLC SSDs RAID-6: | 500 MB/sec writes for each drive<br>22 x 500 = 11000 MB/sec application-level writes at any block size<br>2.75 Million 4K random write IOPS |
| 12 TLC NVMe RAID-5: | 3 GB/sec writes for each drive<br>11 x 3 = 33 GB/sec application-level writes at any block size<br>> 4 Million 4K random write IOPS |

The fastest arrays start to run into other issues that limit IOPS, so ***ESS*** is quoted as > 4M 4K random writes. The actual number depends on your hardware, memory speed, queue depth, memory channels, etc. Even so, actually "needing" 4M random writes is somewhat unusual.

## *ESS* and SSD Wear

When an SSD is used with ***ESS***, it sees only long linear writes. As such, you can use zoned SSDs if available. You can also use lower-cost read-intensive SSDs without being concerned about drive wear.

For example, the QLC Micron 5210 7.68 TB SSD is spec'd as:

0.05 DWPD for 4K random writes

With ***ESS***, this drive will reach about 0.75 DWPD for any workload. This is a QLC drive. TCL drives are in the 2-3 DWPD range for low cost datacenter SSDs. In general, RAID arrays experience about 1/5$^{th}$ or less of the SSD wear that would be experienced for a stock array.

## *ESS* and RAID

Because ***ESS*** controls writes, you can build parity-based arrays instead of using mirrors. Parity-based RAID is faster than the same number of drives mirrored, plus you get more usable capacity.

RAID arrays can also contain very larger numbers of drives. Normally, the thought of a 24 drive SATA array as a single RAID-6 set would be unthinkable. With ***ESS***, this arrays scales perfectly.

## *ESS* Licensing

***ESS*** is used in many products, some without you ever knowing it. WildFire Storage licenses ***ESS*** to storage system vendors with the goal of creating value for the vendor. A storage server built with ***ESS*** can perform better, last longer, but still cost less to build than any competing solution.

### ESS is Stunningly Faster than Standard RAID

Because of how ESS linearizes writes creating the ideal write pattern for both SSDs and arrays, the differential in system and drive overhead is massive. Consider what it takes to support 1 million 4K random writes with a 24 drive RAID-6 array.

| | |
|---|---|
| **Standard RAID 6:** | 21 million 4K random reads<br>these reads are required to compute parity blocks<br>3 million 4K random writes<br>this is the data block plus two parity blocks |
| **ESS RAID 6:** | 140 thousand 32K linear writes<br>this is the 4K data blocks plus parity |

**WildFire Storage**

https://wildfire-storage.com +1 (610) 237-2000
sales@wildfire-storage.com +1 (888) 473-7866