THOUGHT LEADERSHIP

Innovating to Fuel **the Next Decade of Big Data**

by Craig Hibbert

INFINIDAT FEDERAL

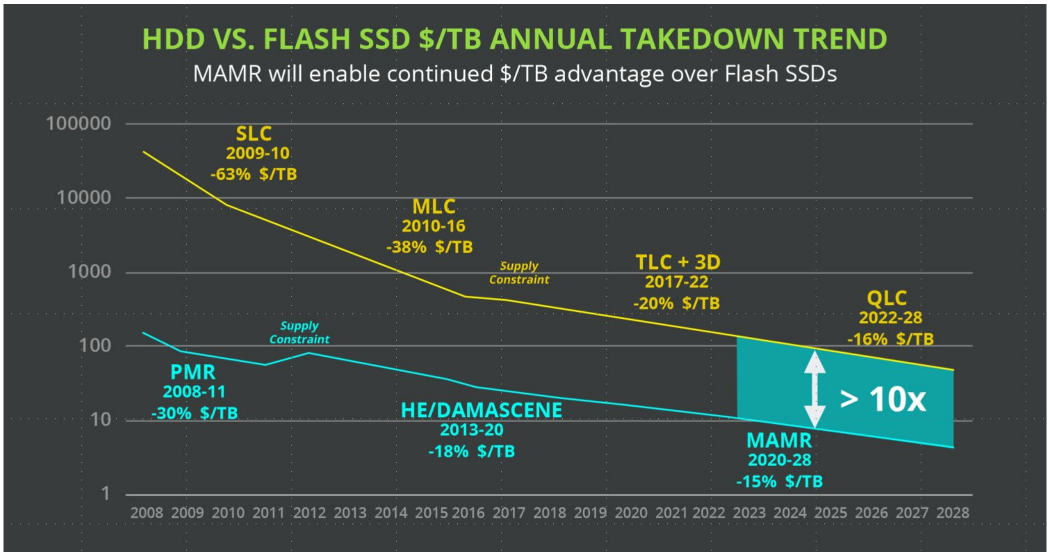
For several decades, storage spend has witnessed huge growth as the adoption of solid-state disks has become the de facto ‘go to’ for most vendors. NAND, as it is often referred to, solves a problem that has plagued the storage array since its inception: “How to handle random reads in an efficient and effective way?” We saw the majority of vendors quickly gravitate to this media as a way of improving overall storage performance without having to spend money innovating themselves. One could argue that this adoption has come at the expense of pioneering new ways to extract additional value from existing technologies that would invariably cost less money, scale higher and have larger sustainable write rates than just a NAND solution.

Technology industry veterans will recognize this as another chapter in the ever-growing book: “Buy the next hardware iteration” to solve all your ailments. Historically, throwing hardware at a solution has only resulted in transferring the targeted problem to somewhere else in the stack. The 10K RPM disk was the gateway to storage freedom. The 15k RPM disk would fix all issues blamed on the (now antiquated) 10K RPM disks. Solid-state was emancipation from spinning disks and NVMe was just the elixir for everything!

Sounds like the wheels of the hardware bus are spinning faster than ever. Of course, these innovations help but there is never one solution to all missions. It is true that vendors who distribute appliances are in some ways reliant upon their hardware manufacturers; but have these technology changes come at the expense of innovation? This document strives to uncover that which is truly different in the storage space - Who does more with less? Who took a different track? What did those internal, introspective and innovative investment cycles mean for the end user/consumer?

Recall from the opening paragraph the benefits that solid-state drives bring with random data. However, the cost point for solid-state enterprise drives is still regarded as being close to 10x the cost of spinning disks, relying on ‘efficacy rates’ to achieve price parity. These technologies are diminished (ergo negated) when end-to-end encryption (the standard in DoD mission environs) is turned on.

Despite some reports showing that the gap is shrinking, spinning drives are still developing new technologies and methodologies that help them continue to expand the gap between themselves and their solid-state brethren. The below table is sourced from a disk manufacturer, arguably the best resource to get next gen drive data from:



Western Digital - Innovating to Fuel the Next Decade of Big Data

The supply side of solid-state media has always lagged behind the demand. Shortages have been exacerbated due to various natural disasters and, more recently, the far-reaching effects of COVID-19. During this timeframe, it has become apparent to various organizations that the sourcing of an alternative solution (outside of NAND) is paramount to averting future potential threats to National Security.

NAND devices can struggle with multiple, disparate workloads (VM's, K8's and Oracle), which drives separation, increased costs and further operational complexity. It was these encumbrances to entire NAND solutions that provided the impetus for the excursion into alternative solutions. Entities sought a strategy to combat price and performance at petabyte-scale and beyond, where the price point was impervious to the use of end-to-end encryption. What if you could address randomness another way? What if you could use spinning drives at significantly lower cost without having to rely on synthetic data reduction savings that evaporate in a fully-encrypted environment? Could you find such a solution? Would it be as resilient as the ones we are accustomed to? Would there be an existing customer bank that could prove its production viability?

Our search was arduous and long, but the answer to the questions posed above is a resounding yes. Not only did we find a vendor operating in this space, but their solution was most intriguing and highly unique when compared against the status quo. It can operate with:

- Only a small percentage of the system being NAND-drive-based
- Equal performance or greater to competitors full NAND systems
- Unique and isolated RAID stripes of data, relevant to the application that wrote the data
- Write out all data sequentially so all subsequent re-reads are sequential
- NL SAS drives (fastest for sequential re-reads than NAND) at a fraction of the cost
- PB scale at highly desired price point with multi protocol File and Block in a unified offering, all equal citizens (meaning one is not operated at the expense of the other)
- All features included (File and Block & immutable Snapshots on both) at one useable price per TB - no extra licensing costs at all
- Next to no maintenance required due to the triple redundancy yielding more 9's of availability than required to protect the nuclear arsenal
- Reduced components, managed automatically itself based upon dynamic metrics on space consumption and resources
- 10 years of history, multiple clients with > 10 frames and multiple PB's deployed, spanning multiple verticals (some we spoke to)

"It is therefore perhaps not surprising that we return to the nucleus of humankind as the next potential breakthrough in recording media."

Our foray into an alternative storage world did not stop with finding unique innovations already deployed and operational. We pushed deeper into 'what comes next' and witnessed some truly remarkable innovations. Humans started cataloging events on stone and (later) parchment. It is therefore perhaps not surprising that we return to the nucleus of humankind as the next potential breakthrough in recording media. It appears that we may be nearing the end of our capability to store data on a two-digit footprint (binary) and need to explore alternative, higher realms with potentially larger orders of magnitude. One such process would use human DNA and nucleotides (bases) to move to a four-value system (Adenine, Cytosine, Guanine and Thiamine). Each human cell has the potential to store slightly less than 2GB of data. The typical human cell contains a genome consisting of about 6 billion base pairs. We challenge our readers to find us a denser storage platform than that!

sales@InfinidatFederal.com | +1.855.893.6994

INFINIDAT FEDERAL