International Journal of Research and Engineering
ISSN: 2348-7860 (O) | 2348-7852 (P) | Vol. 6 No. 2 | March 2019 | PP. 603 - 606
Digital Object Identifier DOI® http://dx.doi.org/10.21276/ijre.2019.6.2.5

Research & Analysis on Hybrid Storage: Combining SSDs and HDDs Drive

ORIGINAL ARTICLE

Author(s): *[1]Muthu Dayalan
Affiliation(s): [1]Senior Software Developer,
ANNA University, Chennai, India
Corresponding author: muthud1985@gmail.com

***Abstract* - The traditional Hard Disk Drives and the upgrading Solid State Drives (SSD) are popular for their storage traits that keep large mass of data. The two technologies provide users with a large number of I/O per second. However, the two have a limited space capacity and perform differently. The SSDs offer exceptional performance. However, as compared to HDDs, they have much less capacity per drive and are costlier. Depending on the organizational application workload, the needs of capacities and performance requirements make users to have different preferences. In the IT system, the challenges on the usage of SSDs and HHDs are navigation of price, performance, and capacity trade-offs between SSDs and HDDs. Determining the most cost-effective drive type for each workload type is essential and will be discussed in this research paper. With the new technologies, there has been a series of combination of SSDs and HDDs into a single storage pool. The reduction of the I/O bottleneck through the hybrid SSD and HDD storage is also analyzed based on the terms of maintaining cost of the cache manager through settling on newer budgets for organizational optimization on storage solutions.**

***Keywords*:** *HHD, SSD, Hybrid storage, workload performance, I/O workload profiles, storage tiering*

I. INTRODUCTION

The traditional mainstay associated with storage technology has always been the hard disk drive (HHD). However, while there is a 40% annual increase in the capacity of HDDs, their input/output (I/O) performance has only increased by 2% annually [1]. This implies that for enterprises that require web, cloud, and virtualized applications may not fully depend on the HDDs for both high capacity and performance. HDDs lack the capabilities to deliver a cost-effective device technology. As a result of these hitch backs, SSDs have gained prominence in the storage solutions industry [1]. SSD offer exceptionally high performance but much less capacity per drive. The cost of SSDs are also considered to be more expensive compared to HDDs and have a write endurance limit [4].

Using the properties of HDDs and SSDs, IT departments can bridge the existing niche on cost effectiveness and high performance through the hybrid storage option [2]. To meet these requirements, integration of HDDs and SSDs into the organizational fabrics is necessary to quantify performance, capacity, and cost value of SSDs. In this research paper, distinct attributes of standalone performance testing and analysis of SSDs and HDDs will account for the different costs, performance, and capacity characteristics [5].

Hybrid hard drives also known as solid-state hybrid drive (SSHD) which is a combination of hard disk drives and solid-state drives, is a combination of mass storage device that combines a conventional hard disk drive and a NAND flash module [3]. It attempts to blend the capacity, cost, and performance of physical disk storage with the accelerated performance of flash.

II. MEASURING HDD AND SSD WORKLOAD PERFORMANCE

In Table 1, application of I/O workload profile is provided to show the two storage devices are used to perform in their different working environment. in an analysis done on the I/O-meter 2006 version that has more randomized data content as compared to the 2008 version.

TABLE 1: APPLICATION I/O WORKLOAD PROFILE

Application	Application I/O Workload Profile			
	Block Size in Bytes	Read/Write Percentage	Random/Sequential Percentage	I/O Performance Metric*
Web File Server	4KB, 8KB, 64KB	95%/5%	75%/25%	IOPS
Database Online Transaction Processing (OLTP)	8KB	70%/30%	100%/0%	IOPS
Exchange Email	4KB	67%/33%	100%/0%	IOPS
OS Drive	8KB	70%/30%	100%/0%	IOPS
Decision Support Systems (DSS)	1MB	100%/0%	100%/0%	IOPS
File Server	8KB	90%/10%	75%/25%	IOPS
Video on Demand	512KB	100%/0%	100%/0%	IOPS
Web Server Logging	8KB	0%/100%	0%/100%	MBPS
SQL Server Logging	64KB	0%/100%	0%/100%	MBPS
OS Paging	64KB	90%/10%	0%/100%	MBPS
Media Streaming	64KB	98%/2%	0%/100%	MBPS
*Metric measurements: IOPS = I/O operations per second MBPS = Megabytes per second				

III. ANATOMY OF A HYBRID HARD DRIVE

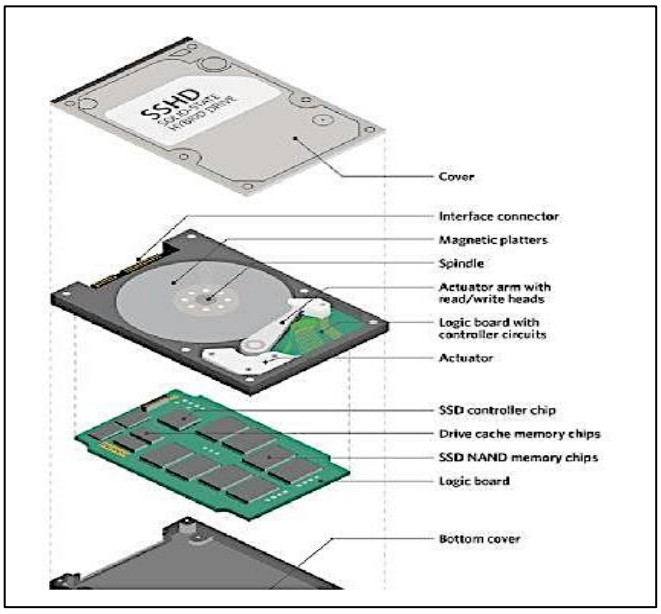


Figure 1: Anatomy of a Hybrid Hard Drive

In relation to the anatomy of HHDs, the HDDs connect to computing devices through a serial-attached SCSI (SAS) or Serial Advanced Technology Attachment (SATA) interface. SSDs on the other hand consists of silicon chips that are designed as an integrated circuit to provide non-volatile memory for resilience. The property makes it distinct from volatile storage on HDDs that require an onboard capacitor or battery backup to protect data in the event of a system failure [6].

A hybrid hard drive adds a small dose of flash to the core disk architecture. A typical HHD cache volume contains about 8 gigabyte (GB) of flash and requires no special software driver. In a research conducted by Kim et al. (2011), hybrid store that integrates the SSDs and HDDs investigates on cost-efficiency and high-performance storage systems. Information technology focuses on storage, protection, retrieval of data in LARGE-SCALE environments [8]. Given the trade-offs between HDDs and SSDs in terms of costs, performance, and lifetime, there is a current consensus among several storage experts to view SSDs not as a replacement of HDD but rather a complimentary device. Within the high-performance storage hierarchy, there is a hybrid system called a high store that provides hybrid plan, HybridDyn. A hybrid plan is used to improve capacity-planning technique among administrators with an overall objective of operating within the cost-budgets. HybridDyn improves performance and lifetime guarantees during episodes of deviations from expected workloads through two novel mechanisms: write-regulation and fragmentation busting. Description of an ideal example of HybridStore's efficacy is a Hybrid Plan that is able to find the most cost-effective storage configuration for a large-scale workload of Microsoft Research.

The Hybrid Dyn under the Microsoft research suggests one MLC SSD with ten 7.2K RPM HDDs instead of the fourteen 7.2K RPM HDDs only. The HybridDyn is able to reduce the average response time for an enterprise scale random-white dominant workload by about 71% as compared to a HDD-based system [9]. The hybrid combination will not only reduce overall power consumption but also enhances responsiveness of a system. In a research on hybrid aggregate project within the Advanced Technology (ATG) exploring on the potential to combine multiple disk types within a single aggregate. The

primary goals of the project is to determine whether a hybrid aggregate that composed of SSDs for cost-effective performance, and Serial-ATA (SATA) disks for their (for their cost-effective capacity), could simultaneously provide better cost/performance and cost/throughout ratios than an all Fiber-Channel (FC) solution.

IV. STORAGE TIERING

Storage tiering increases efficiency by matching business needs the most appropriate storage media. Storage tiering is used by IT organizations by directing I/O operations like OLTP or certain Web-facing applications [10]. Web-facing applications accounts high-performance SSD and Serial attached SCSI (SAS) arrays. An organization that runs a mixture of applications in the course of business operations, each with its own particular set of performance requirements and environmental considerations. Storage architects can configure their PS Series SAN as a heterogeneous storage pool multiple array. The EqualLogic Auto-Tiering features is used to monitor volume size and I/O workloads which then automatically migrate volumes to arrays using the most appropriate disk technologies and RAID policies [13]. Redistribution of workloads across all arrays of a hybrid storage system improves overall SAN performance.

A hybrid drive that has combined SSD and HDD features consists of three storage tiers: SDRAM, NAND memory, and a magnetic disk. Each of the tiers have different write/read speeds. In the case of written data requests, the NAND memory is utilized to store the system requests that data be written. The hybrid drive then directs the written data to the magnetic disk. With the data in the NAND memory, it is easier to send data to the entire system. The integration of SDRAM, NAND memory, and a magnetic disk can also adopt an automatic tiering that moves data based on policies such as data age, frequency of access, and the last time the data is accessed.

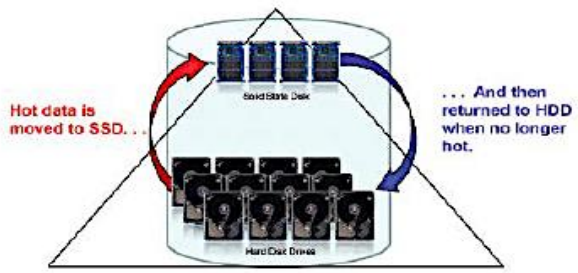
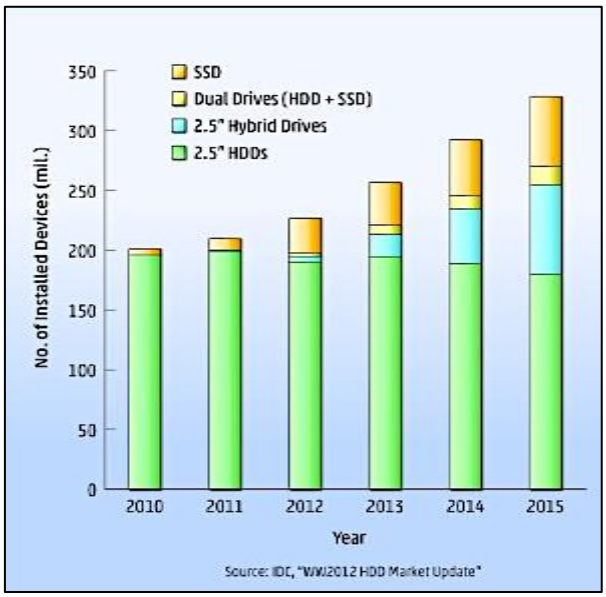


Figure 3: Storage Tiering

V. CURRENT HYBRID DRIVE DESIGNS

The current hybrid drive designs deliver both technologies within a single physical unit. The hybrid drives also employ software caching algorithm to structure data to be stored in the SSD portion and what goes to the HDD platters. System developers design the caching algorithms to reside in the hybrid drive's firmware and not the device driver thus increasing efficiency and data security. The cache is non-volatile thus data does not appear when power is shutdown. Examples of the hybrid designs that have concentrated in the market include the Seagate's Momentus XT SSHD [7]. The designs help the amount of data increasing the tandem with the popularity of sharing data via the internet; the capacity of storage devices such as the HDDs used in PCs is also increasing. The Hybrid drive achieves both a high capacity and high performance by combining traditional NAND memory technologies, NAND memory handling technologies, and HDD product development technologies. In figure 4, a forecast share of built-in storage products for Notebook PCs [14]. A full-fledged adoption of hybrid devices that simultaneously realize both high capacities and high-speed data access performance while keeping bit costs low was observed to increase in 2012 up to the year 2015 with a 25% increase.



Example of a Hybrid Storage - Cloud Storage Application

In a typical IaaS (Infrastructure-as-a-Service), cloud usage is essential in strengthening a system even on temporary basis. Frequent provisioning of virtual volumes is needed. However, the use of a hard disk drive storage is not sufficient thus becoming a bottleneck in the IaaS cloud. A hybrid-storage, distributed storage, and a normal HDD storage are the block storage options of OpenStack Cinder. Hybrid storage achieves a much higher performance compared to the HDD storage making it suitable for analysis applications [14]. The hybrid drive is considered to be cheaper than the solid-state drives since they contain a smaller amount of solid-state memory. With a hybrid drive, both solid-state speeds and mechanical drive storage capacity makes the combination less expensive [12]. A cache algorithm that store operating system and program files in the solid-state memory allows accessibility of the cached files.

VI. CONCLUSIONS

Combination of HDD and SDD storage devices to a hybrid storage has been widely embraced due to the high performance and cost-effectiveness. HDD is considered to have adequate capacity yet their performance is quite slower as compared to the SDD. HDDs are also costly thus making SDD to be preferred by organizations whose needs require a highly performing storage device. The anatomy of a hybrid hard drive clearly demonstrates that there is a strong connection between SSD and HDD in ensuring high performance, and cost-effective option. Storage tiering efficiency is also elaborated on how it makes the hybrid storage to be efficient. The anatomy of a Hybrid drive is defined along with a hybrid system. Within the high-performance storage hierarchy, there is a hybrid system called a high store that provides hybrid plan, HybridDyn. The hybrid drive design differs from one enterprise to another since the needs of owners and clients differ significantly [15]. The caching algorithm is also explained on how it improves performance and accessibility through increased performance and accessibility of files in the solid state. In the IaaS cloud computing system, reliance on the hybrid drive makes it faster and more reliable among users since it is cheaper and faster.

Hybrid storage is a next generation storage due to highly efficient in performance and cost effectiveness.

DECLARATION

All author(s) have disclosed no conflicts of interest. There is no funding to carry out this research and its publication but me as a principle author considers my longstanding interest in enhancing professional career and how that might have affected is the objectivity in research.

REFERENCES

[1] C. King, and T. Vidas, "Empirical analysis of solid-state disk data retention when used with contemporary operating systems." digital investigation 8 (2011): S111-S117.

[2] S. Lee, H. Min, S. Yoon, "Will solid-state drives accelerate your bioinformatics? In-depth profiling, performance analysis and beyond". Briefings in bioinformatics, vol. 17, no.4, pp. 713-727, 2015.

[3] H. Payer, M. Sanvido, Z. Z. Bandic, and C. M. Kirsch, "Combo drive: Optimizing cost and performance in a heterogeneous storage device," in *WISH 2009 The First Workshop on Integrating Solid-state Memory into the Storage Hierarchy* Hamilton Crowne Plaza, Washington DC, USA 2009, pp. 1-8.

[4] J. D. Strunk, "Hybrid Aggregates: Combining SSDs and HDDs in a single storage pool". ACM SIGOPS Operating Systems Review, vol. 46, no. 3, pp. 50-56, 2012.

[5] W. B. Boyle, M. L. Syu, and W. C. Cain, "Hybrid drive comprising write cache spanning non-volatile semiconductor memory and disk." U.S. Patent 8,639,872, January 28, 2014.

[6] N. Alastair, L. Scott, and M. Ruff, "A forensic analysis and comparison of solid-state drive data retention with trim enabled file systems." 2013.

[7] G. McHale, B. G. Nadea, B. K. Panner, P. J. Hunter, D. Hung, & J. Lacy. "U.S. Patent No. 8,775,731. Washington, DC: U.S. Patent and Trademark Office". 2013.

[8] Y. Kim, A. Gupta, B. Urgaonkar, P. Berman, & A. Sivasubramaniam. "HybridStore: A cost-efficient, high-performance storage system combining SSDs and HDDs." In Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on, pp. 227-236. IEEE, 2011.

[9] D. Strunk, "Hybrid Aggregates: Combining SSDs and HDDs in a single storage pool." *A*CM SIGOPS Operating Systems Review, vol. 46, no. 3 (2012): pp. 50-56.

[10] G. Soundararajan, V. Prabhakaran, M. Balakrishnan, and T. Wobber, "Extending SSD Lifetimes with Disk-Based Write Caches." In *FAST*, vol. 10, pp. 101-114. 2010.

[11] Y. Yoji, "Cloud storage application area of HDD–SSD hybrid storage, distributed storage, and HDD storage." *IEEJ Transactions on Electrical and Electronic Engineering* 11, no. 5 (2016): 674-675.

[12] K. Silver, "Realizing SSD Value through Caching Software," in *Flash Memory Summit 2010*, Santa Clara, CA, 2010.

[13] T. Mojtaba, H. Asadi, A. Haghdoost, M. Arjomand, and H. Sarbazi-Azad, "A hybrid non-volatile cache design for solid-state drives using comprehensive I/O characterization." IEEE Transactions on Computers, vol. 65, no. 6, pp. 1678-1691, 2016.

[14] C. Koling, R. Sundaram, D. P. Doucette, and R. Dronamraju, "Hybrid storage aggregate block tracking." U.S. Patent Application 13/413,877, filed September 12, 2013.

[15] K. Youngjae, A. Gupta, B. Urgaonkar, P. Berman, and A. Sivasubramaniam, "HybridStore: A cost-efficient, high-performance storage system combining SSDs and HDDs." In Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on, pp. 227-236. IEEE, 2011.