Review of the Hybrid Storage System

SAMI SARHAN, MAIS HAJ QASEM
Computer Science Department
University of Jordan
Amman, JORDAN
samiserh@ju.edu.jo, mais_hajqasem@hotmail.com

Abstract: - hybrid storage system (HSS) comprises a hierarchy of storage devices that include (from to top to bottom) solid state drive, hard disk drive, and magnetic tape. These storage devices differ from one another in their performance capabilities, particularly in their speed performance, which increases when moving in the hierarchy from the bottom upwards. HSS delivers high performance to business-critical applications while incorporating large-capacity disks to address substantially large data storage repositories. This study discusses a hybrid storage system and presents its importance in various applications. Moreover, this research compares the hybrid storage system levels and presents the importance of migration data in this system.

Keywords: Hybrid Storage System, Solid State Drive, Hard Disk Drive

1 Introduction

A hybrid storage system (HSS) is a term used to describe storage systems that are designed with a blend of flash-based solid-state disk (SSD), hard disk drives (HDD), and magnetic tape to provide high performance at an affordable price compared with storage arrays. However, concerns about hybrid storage include access speed and latency, as well as reservations regarding security, compliance, and data portability [19].

These storage devices differ from one another in their performance capabilities, particularly in their speed performance, which increases when moving in the hierarchy from the bottom upwards [20]. Applications continuously ask for their on-demand data reading requests from the storage system. Therefore, the requested data should be stored in the uppermost level to reduce the retrieving time for such information.

SSD is a component of a hybrid system that receives substantial attention. SSD is at the top of the storage hierarchy, whereas HDD is at the bottom. Thus, SSD is a better data reader in terms of performance with less energy consumption, whereas HDD is better in terms of memory capacity but worst in reading performance [18].

HSSs use flash technology in the form of SSDs to deliver high performance to business-critical applications while incorporating large-capacity disks to address substantially large data storage repositories. Intelligent algorithms within hybrid storage systems move active data to the flash technology tier, thereby enabling lower-accessed data to be stored in low-cost, large-capacity disks [23].

When an application requests data in this hybrid storage system, the system checks its upper-level storage device (i.e., SSD) to locate such data. If the data are not found in SSD, then the system directly checks in the next storage level (i.e., HDD). If the data are also not found in this level, then the system checks the lower level. This process is repeated until the needed data are located. Thus, data should be distributed in the different levels of the hybrid storage system based on their importance [13]. Therefore, hybrid storage systems function by keeping the bulk of data in spinning hard drives and then storing the most frequently used data in SSD for high performance [24].

The value of a hybrid storage system comes from the majority of businesses, where such systems can deliver performance in excess of 20,000 input/output operations per second (IOPs) and as many as 200,000 IOPs. These types of performance may require hundreds to thousands of disks to achieve a legacy disk storage architecture. Hence, the intelligence of the software that use hybrid systems ensure that active data are readily available in the high-performing flash technology layer [22].

The system data reduction capabilities of hybrid systems offer a reduction in the physical space needed to store an effective capacity of data. The function of hybrid storage systems is to store more data on less hardware, thereby providing the ancillary benefit of small footprints, limited power, and small infrastructure to manage [26].

Hybrid systems deliver on the promise of performance without sacrificing availability or data integrity. These systems' turn-key designs include advanced functions, such as data reduction, tiering,

and replication, and are managed through an intuitive user interface, thereby simplifying common user tasks [25].
Hybrid systems enable companies to accomplish substantially with limited effort. Moreover, the use of hybrid systems allows companies to substantially prepare in addressing the unpredictable nature of business and meeting such challenges with existing resources [21].
This study presents an in-depth discussion of HSS and its importance in many applications. In addition, various levels of HSS are compared to one another. The importance of migration of data within a hybrid storage system is likewise discussed.
This paper is organized as follows. Section 2 reviews the related literature on hybrid storage systems. Section 3 presents the importance of prefetching techniques in hybrid storage systems. Section 4 presents a comparison among the various layers of a hybrid storage system. Section 5 presents the conclusion.

2 Hybrid Storage System

An HSS consists of a hierarchy of different storage devices that vary in their speed performance, energy consumption, and size capacity. HSSs provide efficient and low-cost solution to accommodate immense data without affecting the I/O response time. In particular, speed performance and energy consumption are increased as we go down the hierarchy, whereas the size is increased as we go up the hierarchy.
A modern HSS is a typical storage device that includes the main memory, SSD, HDD, and magnetic tape (see illustration in Figure 1). SSD is located at the top of the storage hierarchy followed by HDD at the bottom. Thus, SSD is a better data reader in terms of performance with limited energy consumption, whereas HDD is better in terms of memory capacity but worst in reading performance [19].

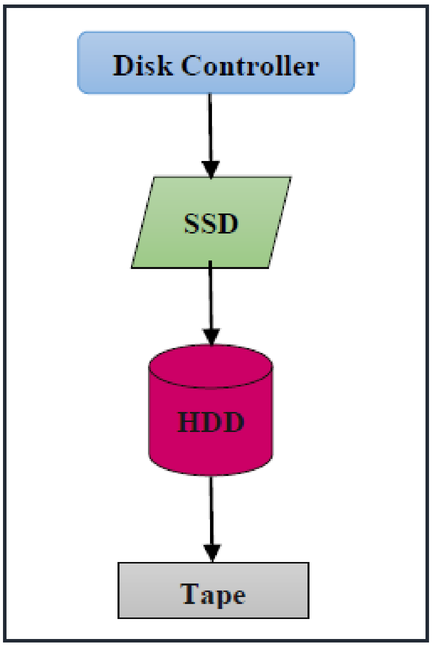

Fig.1 Hybrid Storage System Architecture

When an application requests data, the system checks the upper-level storage device (i.e., SSD) to locate such data. If the data are not found in SSD, then the system directly checks in the next storage level (i.e., HDD). If the data are also not found in this level, then the system checks the lower level. This process is repeated until the needed data are located.

2.1 Parallel Hybrid Storage System

A parallel HSS approach involves a few storage devices, such as SSD, HDD, and tape, to store data that offer high I/O bandwidth and high performance. A parallel storage system is often deployed in super computers. Accordingly, storage devices (e.g., SSD, HDD), each of which is called a disk array, are the most important components. This approach is reliable because it has fault-tolerance features provided by data redundancy [27].
Parallel storage devices provide the I/O parallelism through stripping techniques. Each data block in these techniques is split into small data blocks and stripped through different disk arrays [9]. The disk array is controlled by the disk controller that can respond and coordinate with the request in parallel. Thus, when a few data blocks are requested, the request will be directed by the controller to the multiple disk arrays.

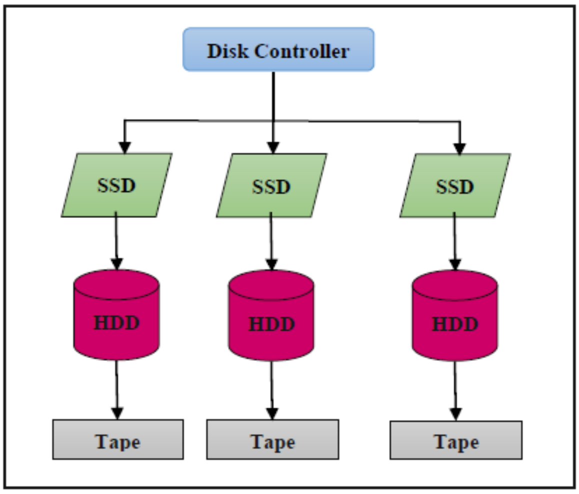

Fig.2 Parallel Hybrid Storage System Architecture

2.2 Solid State Disk (SSD)

SSD is a nonvolatile storage that hold and access persistent data even without power by using flash memory. SSD has no moving parts, thereby enabling it to have a more rapid access time, noiseless operation, higher reliability, and lower power consumption compared with any other storage [28].

SSD has an array of semiconductor memory organized as a disk drive and uses integrated circuits (ICs) rather than magnetic or optical storage media [30]. Internal SSDs connect to a computer by using standard IDE or SATA connections that has a theoretical maximum transfer rate of 750 MB per second [32].

The first SSD was developed by IBM in the 1970s and 1980s for use in their supercomputers. SSD was developed and adopted because of the rapidly expanding need for high I/O performance. Accordingly, SSD is ideal for heavy reading and random workloads because it has considerably lower random access and read access latency than other storage [29].

At present, various hardware, such as high-performance servers, laptops, and desktops, need to deliver information in real-time or near real-time. Hence, this hardware will benefit from the SSD technology. Accordingly, SSD is suitable for offload reads from transaction-heavy databases to alleviate boot storms with virtual desktop infrastructure or inside a storage array to stage hot data locally for off-site storage in a hybrid cloud scenario [31].

2.3 Hard Disk Drive (HDD)

An HDD is a nonvolatile storage that permanently stores and retrieves data from a computer. All data can still be accessed when the computer is turned on. HDD is a secondary storage device that contains magnetic disks or platters rotating at high speeds [34].

An internal HDD is connected to the motherboard using an ATA, SCSI, or SATA cable and powered by a connection to a power supply unit. HDD has many shapes and sizes (i.e., typically 3.5-inch and 2.5-inch models for desktop computers and laptops, respectively). A typical consumer HDD has a rotational speed of 7200 RPM, although a few high-end HDDs run as fast as 15,000 RPM. Laptop HDDs typically run at either 4800 RPM or 5400 RPM [33].

The hard disk was developed by IBM engineers in 1953 to provide random access to high-capacity data at a low cost. The disk drives developed, which could store 3.75 megabytes of data, were the size of refrigerators and were first shipped in 1956 [35]. The first PCs had hard drives that were below 1 megabyte, whereas modern hard drives may contain several terabytes of data. A few desktop computers have multiple internal hard drives, while external hard drives are often used for additional storage or backup purposes [36].

2.4 Magnetic Tape

The magnetic tape is one of the oldest technologies for electronic data storage and used for data collection, backup, and archiving [38]. Tape is made of flexible plastic with one side coated with a ferromagnetic material. Tapes were originally wound in wheel-like reels but were superseded by cartridges and cassettes of many sizes and shapes, thereby providing substantial protection for the tape inside.

The tape remains well-suited for archiving because of its high capacity, low cost, and considerable durability. However, the magnetic tape is a linear recording system that is unideal for random access. If the tape is part of a library, robotic selection and loading of the right cartridge into a tape drive adds substantial latency [37]. In an archive, such latencies are not an issue. However, tape archiving lacks an online copy for rapid retrieval because everything is vaulted for the long term.

2.5 Literature Review on Hybrid Storage System

Table 1 presents various studies that propose parallel hybrid storage systems. Wu and Reddy [2] studied an HSS that employs flash and disk drives. They proposed a measurement-driven migration strategy for managing storage space in such a system to maximize the performance asymmetry of these devices and balance the workload properties across flash and disk drives. This approach automatically and transparently manages the migration of data blocks among flash and disk drives based on their access patterns. The aforementioned researchers showed that the proposed approach is effective based on realistic experiments on a Linux testbed that employed three different benchmarks. The results indicate that the proposed measurement-driven migration can substantially improve the performance of the system by up to 50% in a few cases.

Vinot and Trigui [3] presented the control of a hybrid vehicle using HSS. They also discussed the energy management strategies of such hybrid vehicle, in which two power sources are used: Pontryagin's minimum principle, which involves an optimal energy management strategy; and rule-based parameterized control strategy. The simulation results show that a well-tuned rule-based algorithm presents relatively good performances when compared with the Pontryagin's minimum and remains relevant for various driving cycles. This rule-based algorithm may easily be implemented in a vehicle prototype or in an HIL test bench.

Wu and Reddy [4] considered block allocation and migration to balance the response times of devices across a workload and improve the performance of HSS that employs SSDs and HDDs. They considered the aggregate performance of the devices regardless of the nature (i.e., read/write) and size (i.e., small/large) of the requests. The aforementioned researchers concluded that the proposed approach improved the performance of HSS in terms of enhancing the latency and throughput of the requests.

Lin et al. [5] presented the design, implementation, and evaluation of an HSS called hot random off-loading (HRO). This system is a self-optimizing HSS that uses a fast and small SSD as a by passable cache to hard disks. The goal is to serve a majority of random I/O accesses from the fast SSD based on the history access patterns, particularly the randomness and hotness of the individual files. The researchers developed a 0-1 knapsack optimization model based on the estimates of randomness and hotness to dynamically migrate files between disks and SSD in HRO. Accordingly, the goal is to redirect the hot and random access to SSD. Their experiments demonstrated that small-capacity SSD in HRO improves the overall I/O performance of disks and latency by up to 39% and 23%, respectively.

Vulturescu et al. [6] proposed a hybrid energy source on board electric vehicles. They concentrated on a retrofitting-based design and implemented and tested a hybrid energy storage system on an existent electric microbus. They focused on the implementation and use of the association of high-energy NiCd battery and high-power supercapacitors on board a 3.5 t urban bus. The comparison between a vehicle powered by a hybrid source and the same vehicle powered only by battery highlights the many effects of hybridization. The reduction of losses within the battery is settled by temperature measurements. Moreover, all the highlighted effects show that even for batteries with a good current acceptance, hybridization makes sense at least for its impact on battery losses.

Pascual et al. [7] proposed an energy management strategy for residential microgrids capable of covering all energy demands with an extremely high share of renewable energy sources. There results showed that the combination of electric and thermal storage systems with controllable loads is a promising technology that could maximize the penetration level of renewable energies in the electric system. Designed control substantially improves the power profile exchanged with the grid when compared with the current state-of-the-art strategies in terms of the reduction of power peaks and fluctuations. They concluded from the comparison with a strategy based on a simple moving average. That is, the reductions in the positive and negative power peaks are 52% and 72%, respectively, and the value of THD is reduced by 62%.

Qasem et al. [9] presented a data mining technique on hybrid storage system to classify the data of commercial marketing website that is parallel with the on-demand request. Thereafter, the data were migrated from HDD to SDD based on the expected demand from the data mining techniques that they used. Thereafter, the result of the classification is injected in a parallel hybrid storage system to migrate the data based on their importance. Thus, the uppermost levels accommodate important data. For this purpose, echo state network (ESN), cycle reservoirs with jump (CRJ), and support vector Regression (SVR) are used. They concluded that their proposed approach substantially reduced the elapsed time. Accordingly, the enhancement of the system for the elapsed time was detected to be between 20 % and 36% in their approach.

Xiao et al. [10] proposed a hybrid storage system called performance data synchronization hybrid storage system (PASS) as a tradeoff between the I/O performance and data discrepancy between SSDs and HDDs. PASS includes a high-performance SSD and traditional HDD to store mirrored data for reliability. In this case, all the I/O re-quests are initially redirected to the primary SSD and the updated data blocks are asyn-chronously duplicated thereafter to the backup HDD. The objective of this process is to hide the latency of the duplication operations of the I/O window used to coalesce the write requests and maintain an ordered I/O queue when copying data from the primary SSD to the backup HDD. Their experiments were contacted on the IoMeter, PostMark, and TPCC benchmarks. Moreover, their results showed that PASS can achieve up to 12 times the performance of a RAID1 storage system for the IoMeter and PostMark workloads while tolerating below 2% data discrepancy between the primary SSD and backup HDD.

Table 1. Literature Review of Hybrid Storage System

Paper	Layer	Application	Techniques	Goal
[2]	SSD, HDD	Photovoltaic (PV)-based wireless sensor network	Supercapacitor cells and a Li-ion battery as energy storages	Shows that a combination of both energy storage elements reduces the number of charge/discharge cycle of the Li-ion battery, improving the battery lifetime.
[3]	SSD, HDD	Email server	Combined Flash and disk drives	The presented approach automatically and transparently manages migration of data blocks among flash and disk drives based on their access patterns.
[4]	SSD, HDD	Commercial Netbench	Combined SSD 2.5-inch X25-M 80GB SATA SSD and one 15K RPM with 73G Fujitsu SCSI disk	Designs a space allocation policy that can exploit parallelism across the devices while providing low average latencies.
[5]	SSD, HDD	File system	0-1 knapsack optimization	Reduce the number of long-distance seek operations on disks by designing an allocation algorithm based on the classic 0-1 knapsack optimization problem to dynamically migrate files between disks and the SSD.
[6]	SSD, HDD	Electric vehicles	high energy NiCd battery and high-power supercapacitors on board of a 3.5 t urban bus	Highlighting the effects of the hybridization to reduction of losses within the battery with consequent expected lifetime extension, and improved dynamic of the vehicle and a possible driving range extension.
[7]	SSD, HDD	Residential microgrids capable	Combined electric and thermal storage systems	energy management strategy to covering all the energy demand with a very high share of renewable energy sources
[9]	SSD, HDD	Commercial website	Echo state network and Cycle reservoir with Regular Jump	Migration data of commercial marketing website in parallel with the on-demand request fr
[10]	SSD, HDD	Email, netnews, and e-commerce classes of applications	Combined a high performance SSD and a traditional HDD	Performance data synchronization hybrid storage System (PASS) to tradeoff between I/O performance and data discrepancy between SSDs and HDDs

Table 1 shows that all the researchers focused on two levels of hybrid storage systems (i.e., SSD and HDD) and disregarded the tape level because of its inadequacy, particularly with the emergence of SSD. The literature review agreed with the importance of HSS and its performance in various applications. Moreover, the proposed paper obtained a high result with the use of HSS.

2.1.1 Sub-subsection

When including a sub-subsection you must use, for its heading, small letters, 11pt, left justified, bold, Times New Roman as here.

3 Prefetching in Hybrid Storage System

Prefetching is a well-known technique for solving the I/O bottleneck problems for data-intensive computing [39]. Prefetching techniques have two main types. The first type is a predictive prefetching scheme, which predicts the I/O access type based on the information history for a particular application. The second type is an informed prefetching schema that uses the ability of an application to detect hints for the future I/O access to prefetch the data before such application is submitted [40].

Prefetching in HSS is a solution that aims to reduce the latency of data transforming among SSD, HDD, and tape by caching the most accessed and popular data in SSD. Prefetching techniques in HSS function to predict the midst priority requested data in HDD and the most requested data with high priority in SSD, while leaving the minor requested data in the tape. This approach reduces the application of elapsed time while performing the request [41].

Prefetching techniques are able to classify the data based on their importance and future access. This process enables us to migrate the application data in the HSS levels based on importance. Thus, data are distributed among the different levels based on their importance [13].

The role of prefetching techniques is the classification of data based on their im-portance, thereby generally improving the performance of the system. Data importance variants, by which the data can be classified into groups, include but not limited to data that will be accessed in the near future, most frequently used data, and data that tend to be accessed in a particular event. The intensive application of I/O and their users vary in their access patterns, thereby making data classification depend on the application type. Application types include but are not limited to database applications, image processing applications, marketing, finance, and voice and text recognition, among others. Various classification or data mining approaches may provide better accuracy for a type of application than the rest [18].

Table 2 shows that various prefetching HSS techniques were proposed in the literature. Al Assaf et al. [11] proposed a parallel energy-aware informed prefetching technique called ECO-storage, which makes the application access the most disclosed patterns using the informed prefetching techniques over classes to the data block. Given that SSD is more energy-efficient than HDD, the latter must be standby because power is saved by prefetching the most requested data into the SSD layer.

Al Assaf et al. [12] proposed informed prefetching techniques called IPODS, which prefetch the hinted blocks that makes the application access the most disclosed patterns. This process is implemented using the informed prefetching techniques over to the data block among a distributed multilevel storage system. They developed a prefetching pipeline in IPODS, in which informed prefetching is divided into separate prefetching steps among the multiple level storage in the distributed system. They improved the I/O performance as the data block are prefetched from the hard disk to memory buffer in the remote storage server. The data blocks that are buffered in server prefetched through the network to client's local cache. Subsequently, the pipelining manner improve the I/O performance.

Nijim et al. [13] was the first to propose a multi-layer prefetching algorithm called Pre- HySys, which prefetches data from tape to HDD and from HDD to SSD. The PreHySys algorithm reduces the missing rate of high-end storage components that results in the reduction of the response time for the requested data in HSS.

Jiang et al. [14] proposed a thermal model for HSS that includes HDDs and SSDs. The generated thermal profiles for the HDD and SSD results show that both storage devices are substantially affected by the temperature of the storage node. The two types built

for HSS are called inter-node and intra-node hybrid storage clusters, which estimate the cooling cost of the storage cluster armed with a hybrid storage node. Their thermal model offers two benefits. First, such model enables the reduction of the cost of thermal monitoring. Second, their model enables data center designers to make intelligent decisions related to thermal management during the design stage.

Saha et al. [15] proposed a new data prefetching algorithm called DM-PAS to meet the increasing demand for high-speed data fetching on a large cloud system. The objective of the proposed algorithm is to achieve the energy performance of data prefetching in a hybrid storage system. The simulation shows that the new algorithm offers an improved performance, reliability, and energy efficiency in a multi-layer storage system.

Yoon et al. [8] proposed an intelligent prefetching technique that significantly improves a hybrid flash–disk storage, which is a combination of hard disk and flash memory. They minimized the overall the I/O processing time of HSS using the fully associated sector translation (FAST) technique, which is known as the best mapping method in managing flash memory. The proposed method prefetches objects onto "prefetching" blocks in the file and block levels in HSSs. Their experiments used actual UCC and synthetic data and showed that their proposed prefetching method outperforms the conventional ones.

Zong et al. [16] conducted an in-depth case study on improving the performance of a large-scale HSSs operated by the United States Geological Survey (USGS) Earth Re-sources Observation and Science (EROS) center, which one of the world's largest satellite images distribution systems, through modified cache replacement policy and innovative prefetching algorithms. They used over 5 million real world user download requests provided by EROS to evaluate the effectiveness of the proposed solution. Their results showed that using the least recently used (LRU) caching policy alone are able to achieve an overall 64 % or 70 % hit ratio on a 100 TB or 200 TB FTP servers from composed of SSD respectively. Moreover, the hit ratio can be further improved to 70 % for 100 TB SSDs and 76 % for 200 TB SSDs if intelligent prefetching algorithms are used together with LRU.

Al Assaf et al. [17] presented a predictive prefetching mechanism that is based on the probability graph approach to perform prefetching between various levels in a parallel hybrid storage system. They proposed their approach to invoke the parallel HSS's parallelism and prefetch data among multiple storage levels in parallel with the application's on-demand I/O reading requests. They concluded after they compared their approach with the existing predictive prefetching techniques that the PPHSS approach reduces I/O stalls by retaining data blocks that are predicted to be accessed in the near future cached in the uppermost level of the hybrid storage. Furthermore, PPHSS improves system performance by 4% when using a small cache size and I/O-intensive workload conditions.

Table 2. Prefetching on Hybrid Storage System

Paper	Prefetching Method	HSS Layer	Benchmark	Enhancement
[11]	Informed	SSD, HDD	I/O-intensive	Reduces the power consumption 75 % under I/O-intensive workload conditions
[12]	Informed	SSD, HDD	I/O-intensive	Reduces the total elapsed time by about 6%
[13]	Predictive	SSD, HDD, Tape	Data-intensive	Increased the miss ratio from 75% to 100%.
[14]	Informed	SSD, HDD	Temperature sensors	Minimize negative thermal impacts of hybrid storage clusters.
[15]	Informed	SSD, HDD, Tape	Earth Resources Observation and Science	Offers better performance, reliability, and energy efficiency of hybrid storage systems by 4%
[8]	Informed	SSD, HDD	I/O-intensive	Reduces the total elapsed time by at least 22%
[16]	Informed	SSD, HDD	Earth Resources Observation and Science	improve the FTP server hit ratio to more than 76 %
[17]	Predictive	SSD, HDD	I/O-intensive	Improves system performance by 4%

Table 2 shows that all studies on HSS focus on two factors. Frist, the power consumption of HSS is reduced because SSDs are more energy efficient than HDDs. Thus, prefetching for data in the HDD level to the SDD level enables substantial standby time to save power. Second, the application elapsed time is decreased when prefetching the most requested data and migrating them from HDD to SDD because the latter is faster than the former. Moreover, we observed that the paper that used the informed prefetching provides better performance improvement than the predictive prefetching because of its accurate prefetching decisions.

4 Comparison Between Hybrid Storage Layer

In the last decade, choosing the best storage option required only selecting the highest capacity hard drive that one could afford. However, the last few years have seen a number of new, considerably advanced options arriving in the market and offering several advantages over the old technology [42]. SSDs have gained considerable traction over the last decade and lightning fast performance and durability. Several hybrid storage options have recently emerged as another alternative, thereby allowing users to balance the pros and cons of HDD and SSD technology.

The emergence of SSDs and hybrid drives have significantly altered the storage land-scape, thereby creating a cornucopia of confusing options for the everyday consumer who are faced with the prospect of storing and processing immense quantity of corpo-rate data in the current information-heavy age. Accordingly, consumers should judiciously consider their choice of storage solutions [43].

The main issue for consumers is which type of storage is the better choice. Although no universal answers are provided, several factors should be considered by IT managers before they select which storage type works best for their business, where each business have different needs. Moreover, they have to evaluate the decision based on such needs, preferences, and budget. That is, organizations are likely to base their choice of storage solution on the requirements of their industry as much as the specific advantages and disadvantages of their chosen technology option [50].

SSDs, HDDs, and hybrid storage have the same function: they boot systems and store applications and personal files. However, each type of storage has its own unique feature set. Even though the price of SSDs has been decreasing, the price per gigabyte advantage is still strongly with HDDs [45]. Hybrid storage is the most expensive because they have SSD and HDD in their storage. Moreover, SSDs are more expensive than HDDs in terms of dollar per GB. For the same capacity and form factor, a 1 TB internal 2.5-inch drive translates into 7 cents per gigabyte for HDD and 14 cents per gigabyte for SSD. Given that HDDs are older, more established technologies, they will remain less expensive for the near future. These additional hundreds may push your system price over budget [44].

SSD units above 4 TB are still extremely rare and expensive. In all likelihood, one will find 500 GB to 1 TB units as primary drives in systems, while 500 GB is considered a “base” hard drive in 2015, pricing concerns can push that down to 128 GB for lower-priced SSD-based systems. Multimedia users will require even more with 1 TB to 4 TB drives as common in high-end systems [49].

From the speed side, an SSD-equipped PC will boot in seconds, certainly under a minute, where the hard drive requires time to speed up to operating specs and will continue to be slower than an SSD during normal use [46]. A computer that used an SSD boots faster, launches apps faster, and has faster overall performance. HDD surfaces work best with large files that are laid down in contiguous blocks to enable the drive head to start and end its read in one continuous motion. When hard drives start to fill up, large files can become scattered around the disk platter, which is otherwise known as frag-mentation. Meanwhile, the read/write algorithms have improved to the point that the effect is minimized. However, HDDs can become fragmented, but SSDs do not care where the data are stored on its chips because a physical read head is lacking. Thus, SSDs are inherently faster [48].

SSD is more durable than HDD because the former has no moving parts. Thus, SSD is likely to keep data safe if you drop your laptop bag or your system is shaken by an earthquake while it is operating. However, most hard drives park their read/write heads when the system is off and flying over the drive platter at hundreds of miles an hour when they are in operation. Accordingly, SSD is recommend if the equipment is rough [43].

Product lists in many stores have more models and choice of HDD from different manufacturers for the same capacities than SSDs, thereby making the former more available than SSD [45]. The SSD model lines are increasing, but HDDs are still in the majority for storage devices in PCs. However, HDD relies on spinning platters that make it limited to how small it can be manufactured to make smaller 1.8-inch spinning hard drives, but that's stalled at about 320GB, since the phablet and smartphone manufacturers have settled on flash memory for their primary storage. SSDs have no such limitation, thereby enabling the continuous manufacture of smaller versions over time [47].

Table 3. Comparison Between Hybrid Storage System Layer

Properties	SSD	HDD
Battery Life	• Less power • Averages 2 – 3 watts • Resulting in 30+ minute battery boost	• More power draws • Averages 6 – 7 watts • Uses more battery
Cost	Expensive	very cheap
Capacity	1TB max for notebook, and 4TB max for desktop	500GB- 2TB from notebook, and 10TB max for desktop
Operating System Boot Time	10-13 second s	30-40 seconds
Magnetism Affected	Safe	Erase Data
File Opening Speed	30% faster than HDD	Slower than SSD
Vibration	No	Yes
Heat Produced	little heat is produced	more heat than an SSD
File Copy / Write Speed	200 MB/s - 550 MB/s	50 MB / s – 120 MB / s
Encryption	Full Disk Encryption	Full Disk Encryption
Weight	SSD drives are lighter than HDD	HDDs are heavier than SSD drives
Operating System	Requires min Win 7 or Mac OS X for good and better performance	All OS

Therefore, HDDs is advantageous in terms of price, capacity, and availability. How-ever, SSDs work best if speed, ruggedness, form factor, and fragmentation are important factors. Moreover, SSD is the choice if performance and fast boot up are the primary considerations and money is secondary. Table 3 illustrates the comparison be-tween SSD and HDD.

4 Conclusion

Hybrid storage system consists of a hierarchy of storage devices, which include, from to top to A hybrid storage system comprises a hierarchy of storage devices, including (from top to bottom) SSD, HDD, and tape. These storage devices differ from one another in terms of their performance capabilities, particularly their speed performance, which increases when moving in the hierarchy from the bottom upwards.

Researchers that conduct studies on hybrid storage systems focus on two factors. First, they endeavor to reduce the power consumption of HSS because SSDs are more energy efficient than HDDs. Accordingly, prefetching data from the HDD to SDD levels enables it to have as much standby time as possible to save power. Second, researchers aim to decrease the application elapsed time when prefetching the most requested data and migrate it from HDD to SDD because the latter is faster than the former. In addition, researchers focus exclusively focus on SSD and HDD, thereby disregarding the tape level because of its limitations, particularly with the emergence of SSD. The literature review agreed with the importance of HSS and its high-quality performance in various applications.

References:

[1] Ongaro, F., Saggini, S., & Mattavelli, P. (2012). Li-ion battery-supercapacitor hybrid storage system for a long lifetime, photovoltaic-based wireless sensor network. IEEE Transactions on Power Electronics, 27(9), 3944-3952.

[2] Wu, X., & Reddy, A. N. (2009, September). Managing storage space in a flash and disk hybrid storage system. In Modeling, Analysis & Simulation of Computer and Telecommunication Systems, 2009. MASCOTS'09. IEEE International Symposium on (pp. 1-4). IEEE.

[3] Vinot, E., & Trigui, R. (2013). Optimal energy management of HEVs with hybrid storage system. Energy Conversion and Management, 76, 437-452.

[4] Wu, X., & Reddy, A. N. (2010, August). Exploiting concurrency to improve latency and through-put in a hybrid storage system. In Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2010 IEEE International Symposium on (pp. 14-23). IEEE.

[5] Lin, L., Zhu, Y., Yue, J., Cai, Z., & Segee, B. (2011, July). Hot random off-loading: A hybrid storage system with dynamic data migration. In Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on (pp. 318-325). IEEE.

[6] Vulturescu, B., Trigui, R., Lallemand, R., & Coquery, G. (2013). Implementation and test of a hybrid storage system on an electric urban bus. Transportation Research Part C: Emerging Technologies, 30, 55-66.

[7] Pascual, J., Sanchis, P., & Marroyo, L. (2014). Implementation and control of a residential electrothermal microgrid based on renewable energies, a hybrid storage system and demand side management. Energies, 7(1), 210-237.

[8] Yoon, U. K., Kim, H. J., & Chang, J. Y. (2010, August). Intelligent data prefetching for hybrid flash-disk storage using sequential pattern mining technique. In Computer and Information Science (ICIS), 2010 IEEE/ACIS 9th International Conference on (pp. 280-285). IEEE.

[9] Qasem, M. H., Al Assaf, M. M., & Rodan, A. (2015). Data mining approach for commercial data classification and migration in hybrid storage systems (Doctoral dissertation, The University of Jordan).

[10] Xiao, W., Lei, X., Li, R., Park, N., & Lilja, D. J. (2012, July). Pass: a hybrid storage system for performance-synchronization tradeoffs using ssds. In Parallel and Distributed Processing with Applications (ISPA), 2012 IEEE 10th International Symposium on (pp. 403-410). IEEE.

[11] Al Assaf, M. M., Jiang, X., Abid, M. R., & Qin, X. (2013). Eco-storage: A hybrid storage system with energy-efficient informed prefetching. Journal of Signal Processing Systems, 72(3), 165-180.

[12] Al Assaf, M. M., Alghamdi, M. I., Jiang, X., Zhang, J., & Qin, X. (2012, August). A pipelining approach to informed prefetching in distributed multi-level storage systems. In Network Compu-ting and Applications (NCA), 2012 11th IEEE International Symposium on (pp. 87-95). IEEE.

[13] Nijim, M., Zong, Z., Qin, X., & Nijim, Y. (2010, September). Multi-layer prefetching for hybrid storage systems: algorithms, models, and evaluations. In Parallel Processing Workshops (ICPPW), 2010 39th

International Conference on (pp. 44-49). IEEE.
[14] Jiang, X., Al Assaf, M. M., Zhang, J., Alghamdi, M. I., Ruan, X., Muzaffar, T., & Qin, X. (2013). Thermal modeling of hybrid storage clusters. Journal of Signal Processing Systems, 72(3), 181-196.
[15] Saha, S., Biswas, A., Nijim, M., & McLauchlan, L. (2014, April). Energy Efficiency Evaluation of a Data Mining Prefetching Algorithm for Hybrid Storage Systems. In Green Technologies Conference (GreenTech), 2014 Sixth Annual IEEE (pp. 99-103). IEEE.
[16] Zong, Z., Fares, R., Romoser, B., & Wood, J. (2014). FastStor: improving the performance of a large scale hybrid storage system via caching and prefetching. Cluster Computing, 17(2), 593-604.
[17] Al Assaf, M. M. (2015). Predictive Prefetching for Parallel Hybrid Storage Systems. International Journal of Communications, Network and System Sciences, 8(05), 161.
[18] Nijim, M., Nijim, Y., Sker, R., Reddy, V., & Raju, R. N. (2011, September). DM-pas: A data mining prefetching algorithm for storage system. In High Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on (pp. 500-505). IEEE.
[19] Levy, R. S. (2012). U.S. Patent No. 8,315,995. Washington, DC: U.S. Patent and Trademark Of-fice.
[20] Cher, J., Kufner, A. C., Sengebusch, F., Chouthe, S., Wang, X., & Li, Y. (2017). U.S. Patent Application No. 15/123,158.
[21] 21. Zhang, G., Tang, X., & Qi, Z. (2010). Application of hybrid energy storage system of super-capacitors and batteries in a microgrid. Dianli Xitong Zidonghua(Automation of Electric Power Systems), 34(12), 85-89.
[22] Oh, Y., Choi, J., Lee, D., & Noh, S. H. (2012, February). Caching less for better performance: balancing cache size and update cost of flash memory cache in hybrid storage systems. In FAST(Vol. 12).
[23] Prabhakaran, V., Balakrishnan, M., & Soundararajan, G. (2013). U.S. Patent No. 8,407,403. Washington, DC: U.S. Patent and Trademark Office.
[24] Lee, H. G. (2010). High-performance NAND and PRAM hybrid storage design for consumer electronics. IEEE Transactions on Consumer Electronics, 56(1).
[25] de Castro, R., Pinto, C., Araújo, R. E., Melo, P., & Freitas, D. (2012, October). Optimal sizing and energy management of hybrid storage systems. In Vehicle Power and Propulsion Conference (VPPC), 2012 IEEE (pp. 321-326). IEEE.
[26] Karden, E., Ploumen, S., Fricke, B., Miller, T., & Snyder, K. (2007). Energy storage devices for future hybrid electric vehicles. Journal of Power Sources, 168(1), 2-11.
[27] Cao, J., & Emadi, A. (2012). A new battery/ultracapacitor hybrid energy storage system for elec-tric, hybrid, and plug-in hybrid electric vehicles. IEEE Transactions on power electronics, 27(1), 122-132.
[28] Kim, J., Hong, A. J., Kim, S. M., Song, E. B., Park, J. H., Han, J., ... & Wang, K. L. (2009, June). Novel Vertical-Stacked-Array-Transistor (VSAT) for ultra-high-density and cost-effective NAND Flash memory devices and SSD (Solid State Drive). In VLSI Technology, 2009 Sympo-sium on (pp. 186-187). IEEE.
[29] Steele, S. A. (2017). Solid State Drive.
[30] 30. Jahns, T. M. (1980). Improved reliability in solid-state AC drives by means of multiple independ-ent phase drive units. IEEE Transactions on Industry Applications, (3), 321-331.
[31] Jahns, T. M. (1980). Improved reliability in solid-state AC drives by means of multiple independ-ent phase drive units. IEEE Transactions on Industry Applications, (3), 321-331.
[32] Heo, J. G., Lee, D., Hwang, S., Lee, D., Lee, J. W., Cheon, W., ... & Jee, W. (2012). U.S. Patent No. 8,327,066. Washington, DC: U.S. Patent and Trademark Office.
[33] Chen, B. M., Lee, T. H., Venkataramanan, V., & Chen, B. M. (2002). Hard disk drive servo sys-tems (p. 47). Berlin: Springer.
[34] Wood, R. (2009). Future hard disk drive systems. Journal of magnetism and magnetic materi-als, 321(6), 555-561.
[35] Chen, Y., Moore, K. L., Yu, J., & Zhang, T. (2006, December). Iterative learning control and repetitive control in hard disk drive industry-a tutorial. In Decision and Control, 2006 45th IEEE Conference on (pp. 2338-2351). IEEE.

[36] Al Mamun, A., Guo, G., & Bi, C. (2006). Hard disk drive: mechatronics and control. CRC press.

[37] Wilkinson, R. T., & Houghton, D. (1975). Portable four-choice reaction time test with magnetic tape memory. Behavior Research Methods & Instrumentation, 7(5), 441-446.

[38] Brady, M. J., Duan, D. W., & Kodukula, V. S. (2001). U.S. Patent No. 6,201,474. Washington, DC: U.S. Patent and Trademark Office.

[39] Kotz, D., & Ellis, C. S. (1991, December). Practical prefetching techniques for parallel file sys-tems. In Parallel and Distributed Information Systems, 1991., Proceedings of the First Interna-tional Conference on (pp. 182-189). IEEE.

[40] Vander Wiel, S. P., & Lilja, D. J. (1997). When caches aren't enough: Data prefetching tech-niques. Computer, 30(7), 23-30.

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

[42] Rizvi, S. S., & Chung, T. S. (2010, April). Flash SSD vs HDD: High performance oriented modern embedded and multimedia storage systems. In Computer Engineering and Technology (ICCET), 2010 2nd International Conference on (Vol. 7, pp. V7-297). IEEE.

[43] Welch, B., & Noer, G. (2013, May). Optimizing a hybrid SSD/HDD HPC storage system based on file size distributions. In Mass Storage Systems and Technologies (MSST), 2013 IEEE 29th Symposium on (pp. 1-12). IEEE.

[44] Kasavajhala, V. (2011). Solid state drive vs. hard disk drive price and performance study. Proc. Dell Tech. White Paper, 8-9.

[45] Wong, G. (2013). SSD market overview. In Inside Solid State Drives (SSDs) (pp. 1-17). Springer, Dordrecht.

[46] Zhang, Y., & Swanson, S. (2015, May). A study of application performance with non-volatile main memory. In Mass Storage Systems and Technologies (MSST), 2015 31st Symposium on(pp. 1-10). IEEE.

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

[48] Yang, P. Y., Jin, P. Q., & Yue, L. H. (2012). A time-sensitive and efficient hybrid storage model involving SSD and HDD. Jisuanji Xuebao(Chinese Journal of Computers), 35(11), 2294-2305.

[49] Domingo, J. S. (2014). SSD vs. HDD: what's the difference. PC Magazine.

[50] Yang, P., Jin, P., Wan, S., & Yue, L. (2013, June). HB-Storage: optimizing SSDs with a HDD write buffer. In International Conference on Web-Age Information Management (pp. 28-39). Springer, Berlin, Heidelberg.