US 20160034202A1

(19) **United States**

(12) **Patent Application Publication**

ROMEM et al.

(10) **Pub. No.: US 2016/0034202 A1**

(43) **Pub. Date: Feb. 4, 2016**

(54) **MULTI-TIERED STORAGE DEVICE SYSTEM AND METHOD THEREOF**

(71) Applicant: **EXCELERO**, Tel Aviv (IL)

(72) Inventors: **Yaniv ROMEM**, Jerusalem (IL); **Omri MANN**, Jerusalem (IL); **Ofer OSHRI**, Kfar Saba (IL)

(73) Assignee: **EXCELERO**, Tel Aviv (IL)

(21) Appl. No.: **14/746,878**

(22) Filed: **Jun. 23, 2015**

Related U.S. Application Data

(60) Provisional application No. 62/126,920, filed on Mar. 2, 2015, provisional application No. 62/119,412, filed on Feb. 23, 2015, provisional application No. 62/096,908, filed on Dec. 26, 2014, provisional application No. 62/085,568, filed on Nov. 30, 2014, provisional application No. 62/030,700, filed on Jul. 30, 2014.

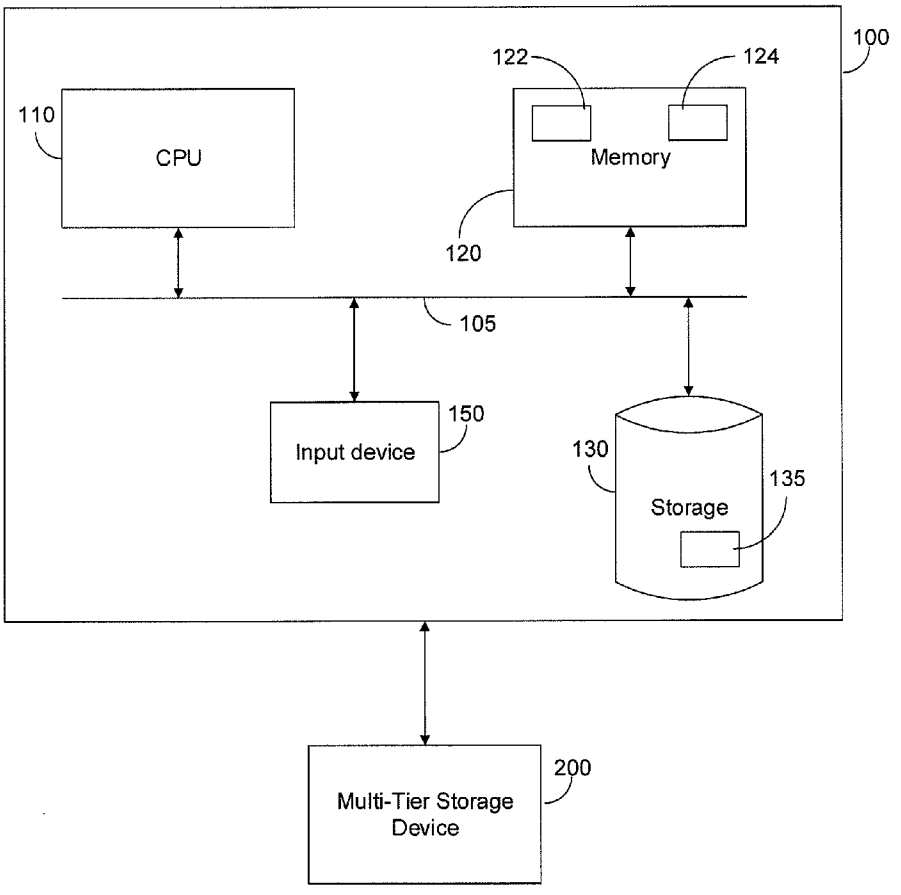
Publication Classification

(51) **Int. Cl.**
G06F 3/06 (2006.01)

(52) **U.S. Cl.**
CPC ***G06F 3/0611*** (2013.01); ***G06F 3/065*** (2013.01); ***G06F 3/0647*** (2013.01); ***G06F 3/0659*** (2013.01); ***G06F 3/0688*** (2013.01); ***G06F 3/0689*** (2013.01); ***G06F 3/0652*** (2013.01)

(57) **ABSTRACT**

A system for controlling a multi-tiered storage device (MTSD) includes a first group of storage devices and a second group of storage devices, wherein each storage device of the first group of storage devices has a higher endurance than each storage device of a second group of storage devices of the MTSD. A block of data is received by a controller of the MTSD, and is written to a storage device of the first group of storage devices. Upon determination that the block of data has been written to infrequently, the block of data is written to a storage device of the second group of storage devices. The block of data may then be erased from the storage device of the first group of storage devices. In some embodiments, a storage device from the first group is associated with the second group, upon determination that the storage has degraded.



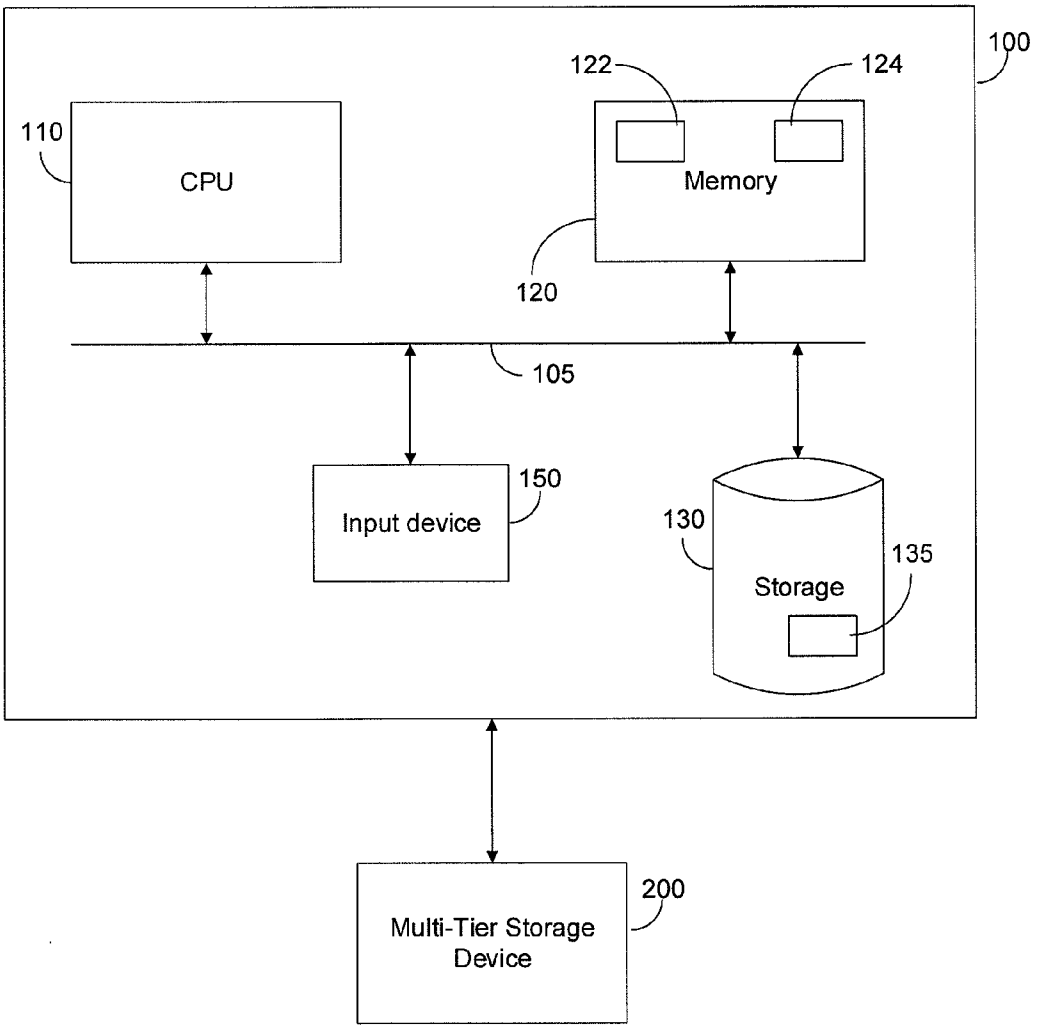

100
122
124
110
CPU
Memory
120
105
150
Input device
130
135
Storage
200
Multi-Tier Storage Device

FIG. 1

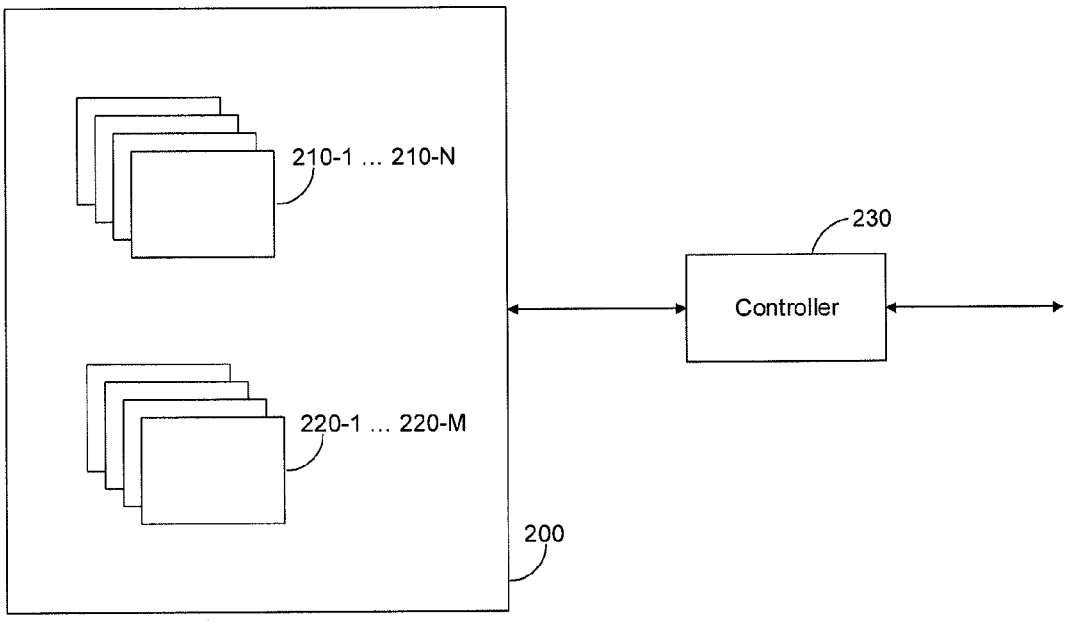


FIG. 2

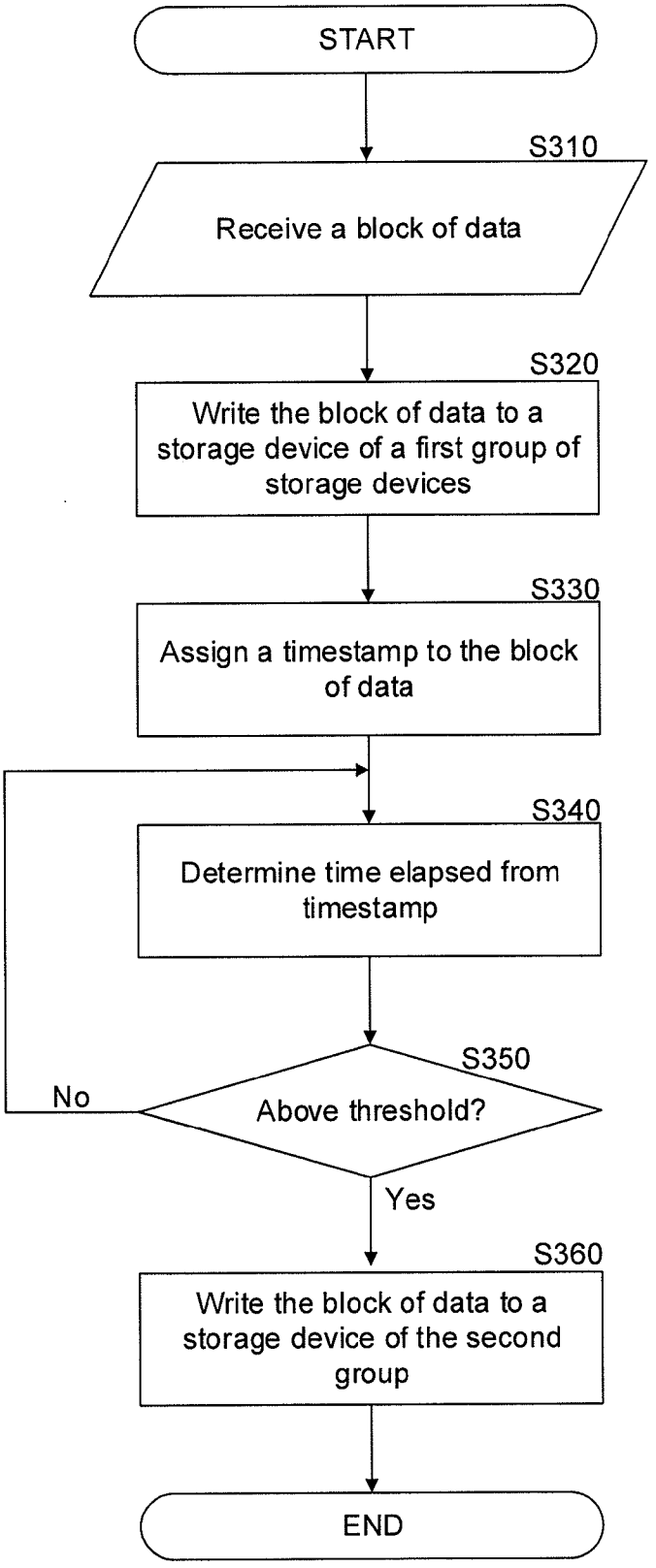

START
S310
Receive a block of data
S320
Write the block of data to a storage device of a first group of storage devices
S330
Assign a timestamp to the block of data
S340
Determine time elapsed from timestamp
S350
Above threshold?
No
Yes
S360
Write the block of data to a storage device of the second group
END

FIG. 3

MULTI-TIERED STORAGE DEVICE SYSTEM AND METHOD THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This Application claims the benefit of U.S. provisional Application Nos. 62/126,920 filed on Mar. 2, 2015, 62/119,412 filed on Feb. 23, 2015, 62/096,908 filed on Dec. 26, 2014, 62/085,568 filed on Nov. 30, 2014, and 62/030,700 filed on Jul. 30, 2014, the entire disclosures of which are incorporated herein by reference for all purposes.

BACKGROUND

[0002] 1. Field

[0003] The disclosure generally relates to multi-tiered storage devices and particularly to controlling multi-tiered storage devices.

[0004] 2. Description of Related Art

[0005] Storage devices constructed in accordance with different technologies offer different advantages for different types of data. Hard disk drives (HDDs) are typically electromechanic devices storing data magnetically, offering low-cost storage. Solid-state drives (SSDs) are storage devices based on integrated circuits (ICs), using for example flash memory and offering higher-cost storage. While byte-for-byte SSDs are more expensive than HDDs, an SSD offers lower access times and lower latency than HDDs. Solutions exist which attempt to consolidate the advantages of different technologies, for example by creating hybrid drives, however the efficiency of these solutions is varying. Furthermore, SSDs constructed in accordance with different technologies offer different levels of write endurance.

SUMMARY

[0006] Exemplary embodiments overcome the above disadvantages and other disadvantages not described above. Also, an exemplary embodiment is not required to overcome the disadvantages described above, and an exemplary embodiment of the present inventive concept may not overcome any of the problems described above.

[0007] One or more exemplary embodiments provide a method for controlling a multi-tiered storage device (MTSD). The method includes receiving a block of data using a controller of the MTSD, the MTSD including a first storage including one or more storage devices, and a second storage including one or more storage devices, wherein each storage device of the first storage has a higher write endurance than each storage device of the second storage; writing, by the controller, the received block of data to a first storage device of the first storage; assigning a timestamp respective of the received block of data; writing, by the controller, the received block of data to a second storage device of the second storage, in response to determining that a time elapsed since the timestamp is above a first threshold value.

[0008] The method may further include erasing the received block of data from the first storage device of the first storage, in response to determining that the received block of data was written to the second storage device of the second storage.

[0009] A storage device of the first storage may be one from among a solid-state drive (SSD), and a hard-disk drive (HDD).

[0010] The method may further include detecting a degradation in a physical parameter of the first storage device of the first storage; and associating the first storage device with the second storage.

[0011] The method may further include duplicating at least a portion of data stored on the first storage device of the first storage to another storage device of the first storage; and removing the duplicated data from the first storage device of the first storage.

[0012] Detection of the degradation in the physical parameter may include determining that the physical parameter of the first storage device is below a second threshold value, the physical parameter includes at least one of endurance, performance, update latency and access time.

[0013] According to an aspect of another exemplary embodiment there is a system for controlling a multi-tiered storage device (MTSD). The system includes a processing unit (PU); an input/output (I/O) interface communicatively coupled to the MTSD, the MTSD includes a first storage including one or more storage devices, and a second storage including one or more storage devices, each storage device of the first storage has a higher write endurance than each storage device of the second storage; and a memory communicatively coupled to the PU, the memory storing instructions that when executed by the PU configure the system to: receive a block of data; write the received block of data to a first storage device of the first storage; assign a timestamp respective of the received block of data; and write the received block of data to a second storage device of the second storage, upon determination that a time elapsed since the timestamp is above a first threshold value.

[0014] The memory may further store instructions that when executed by the PU configure the system to erase the received block of data from the first storage device of the first storage, in response to determining that the received block of data was written to the second storage device of the second storage.

[0015] The first storage device may be one from among a solid-state drive (SSD), and a hard-disk drive (HDD).

[0016] The memory may further store instructions that when executed by the PU configure the system to: detect a degradation in a physical parameter of the first storage device of the first storage; and associate the first storage device with the second storage.

[0017] The memory may further store instructions that when executed by the PU configure the system to: duplicate at least a portion of data stored on the first storage device of the first storage to another storage device of the first storage; and remove the duplicated data from the first storage device.

[0018] The memory may further store instructions that when executed by the PU to detect the degradation in the physical parameter further configure the system to determine that the physical parameter of the first storage device is below a second threshold value, the physical parameter includes at least one of endurance, performance, update latency and access time.

[0019] A non-transitory computer readable medium having stored thereon instructions for causing one or more processing units to execute any of the methods described above.

[0020] According to another exemplary aspect of the present disclosure there is a system including: a controller; and a memory that stores information arranging storage devices in a multi-tiered storage device (MTSD) into at least two groups including a first group of storage devices and a

second group of storage devices, the storage devices in the first group of storage devices have a write endurance that is higher than a write endurance of the storage devices of the second group of storage devices, wherein in response to the write endurance of a first storage device of the first group of storage devices falling below a threshold, the controller re-assigns the first storage device to the second group of storage devices.

[0021] In response to the write endurance of a first storage device of the first group of storage devices falling below a threshold, the controller may further cause data on the first storage device to be written to a second storage device of the first group of storage devices.

[0022] The system may be connected to the MTSD over a network.

[0023] Other features and aspects will be apparent from the following detailed description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] The foregoing and other objects, features and advantages will become apparent and more readily appreciated from the following detailed description taken in conjunction with the accompanying drawings, in which:

[0025] FIG. **1** is a schematic illustration of system for controlling a multi-tiered storage device (MTSD) implemented according to an embodiment.

[0026] FIG. **2** is a schematic illustration of an MTSD in accordance with an embodiment.

[0027] FIG. **3** is a flowchart of a method for controlling an MTSD in accordance with an embodiment.

DETAILED DESCRIPTION

[0028] Below, exemplary embodiments will be described in detail with reference to accompanying drawings so as to be easily realized by a person having ordinary skill in the art. The exemplary embodiments may be embodied in various forms without being limited to the exemplary embodiments set forth herein. Descriptions of well-known parts are omitted for clarity, and like reference numerals refer to like elements throughout.

[0029] The exemplary embodiments disclosed herein are only examples of the many advantageous uses of the innovative teachings herein. In general, statements made in the specification of the present application do not necessarily limit any of the various claims. Moreover, some statements may apply to some inventive features but not to others. In general, unless otherwise indicated, singular elements may be in plural and vice versa with no loss of generality.

[0030] Storage devices degrade over time. For example, storage devices have an endurance. The term endurance as used herein can be considered as any of the life span of the storage device that a user can expect when reading and writing data, changing performance of the storage device over time, changing update latency of the storage device over time, changing access time of the storage device over time, and the like. A multi-tiered storage device (MTSD) includes a first group of storage devices and a second group of storage devices. The groups of storage devices may be determined by the relative endurance of the storage devices. For example, each storage device of the first group of storage devices may have a higher endurance than each storage device of a second group of storage devices of the MTSD.

[0031] A block of data may be received by a controller of the MTSD, and written to a first storage device that is part of the first group of storage devices. Upon determination that the block of data has been written to infrequently, the block of data may be written to a second storage device that is part of the second group of storage devices and the block of data may be erased from the first storage device. In this way, the life span of the second storage device (or any storage device in the second group of storage devices) may be preserved. In some exemplary embodiments, the classification of a storage device may change from the first group of storage devices to the second group of storage device. For example, over time the endurance of a certain storage device may degrade. In this case, that certain storage device, which was originally part of the first group of storage devices, becomes part of the second group of storage devices. This reclassification may occur in response to determining that the endurance or lifespan of a certain storage device has degraded beyond the threshold for being classified in the first group. The threshold for determining which classification a storage device should receive is arbitrary and may be set by the designer.

[0032] FIG. **1** is an exemplary and non-limiting schematic illustration of a system **100** for controlling a multi-tiered storage device implemented according to an exemplary embodiment. The system **100** includes at least one processing element **110**, for example, a central processing unit (CPU). The CPU is coupled via a bus **105** to a memory **120**. The memory **120** further includes a memory portion **122** that contains instructions that when executed by the processing element **110** perform the method described in more detail herein. The memory **120** may be further used as a working scratch pad for the processing element **110**, a temporary storage, and others, as the case may be. The memory **120** may include volatile memory such as, but not limited to, random access memory (RAM), or non-volatile memory (NVM), such as, but not limited to, Flash memory. The memory **120** may further include memory portion **124** containing a timestamp respective of received blocks of data. The processing element **110** may be coupled to an input device **150** via the bus **105**, for example. The processing element **110** may be further coupled with a database **130** via the bus **105**, for example. The database **130** may hold a copy of the method executed in accordance with the disclosed technique. The database **130** may further include storage portion **135** containing a list of storage devices associated with an MTSD **200**. The system **100** is communicatively coupled with the MTSD **200**.

[0033] In some exemplary embodiments, the MTSD **200** is coupled via the bus **105** to the CPU. In other exemplary embodiments, the MTSD may be communicatively coupled with the system **100** through, for example, a network. The network may be configured to provide connectivity of various sorts, as may be desired based on the particular implementation, including but not limited to, wired and/or wireless connectivity, including, for example, local area network (LAN), wide area network (WAN), metro area network (MAN), worldwide web (WWW), Internet, and any combination thereof, as well as cellular connectivity.

[0034] FIG. **2** is a non-limiting exemplary schematic illustration of the MTSD **200**. The MTSD **200** includes a first group of storage devices **210-1** through **210**-N and a second group of storage devices **220-1** through **220**-M, where 'N' and 'M' are integers having a value equal to, or greater than '1'. N and M may be the same value or different values. The first group of storage devices **210** is characterized in that each

storage device **210-1** through **210**-N has a higher write endurance than each of the storage devices of the second group of storage devices **220**. For example, an MTSD **200** may associate all storage devices with a write endurance of fifty thousand (50,000) cycles and above with the first group of storage devices, and all storage devices with a write endurance below 50,000 cycles to the second group of storage devices **220**. The MTSD **200** is coupled with a controller **230**. In some embodiments, the first group of storage devices **210** may include solid state drives (SSDs). In certain embodiments, the second group of storage devices **220** may include SSDs and/or hard disk drives (HDDs). In some embodiments, the MTSD may have more than two tiers (groups) of storage devices, wherein each tier includes storage devices which each have a write endurance level that is within the threshold of that tier. Some storage technologies, such as solid-state drive technology, have a natural degradation of the storage device. In such embodiments, a controller may re-designate a degraded storage device from the first group of storage devices **210** to the second group of storage devices **220**. Degradation may be determined, for example, by a controller **230** or a system **100**, according to a threshold respective of a physical characteristic of a storage device, such as access time of the device, update latency, endurance or performance.

[0035] FIG. **3** is a non-limiting exemplary flowchart of a method for controlling a multi-tiered storage device (MTSD) in accordance with an embodiment. In S**310** a block of data is received by a controller **230** to be written to the MTSD **200**. In S**320** the received block of data is written by the controller **230** to a storage device of a first group of storage devices **210** of the MTSD **200**. Each storage device of the first group of storage devices **210** has a higher endurance than each storage device of a second group of storage devices **220** of the MTSD **200**. In S**330** a timestamp is assigned to the block of data, to determine the time in which the received block of data was written. In S**340** a time elapsed is determined from the timestamp. In S**350** a check is performed to determine if the determined time elapsed since the timestamp is above a determined first threshold. If yes, execution continues at S**360**, otherwise execution continues at S**340**. The threshold may be static, dynamic or adaptive. Static thresholds are predetermined thresholds that remain constant. Dynamic thresholds are forcefully changed, for example, at a certain time of day, a certain day of the year, and the like. Adaptive thresholds are changed in response to changes in characteristics of the MTSD **200** and may vary depending on a variety of parameters, such as total free memory in the MTSD **200**, and the like. In **5360** the received block of data is written to a storage device of the second group of storage devices **220**. The block of data on the storage device of the first group of storage devices **210** is now redundant, and may be erased, in some embodiments.

[0036] In some embodiments, a degradation in a physical parameter of a storage device **210-1** of the first group of storage devices **210** is detected. Detection of the degradation includes determining that the physical parameter of the at least a storage device is below a second threshold value. The physical parameter may be endurance, performance, update latency, access time, and the like. At least a portion of data stored on the storage device **210-1** is duplicated and stored on the storage device **210-2** of the first group of storage devices **210**. Once the data is stored on the second storage device **210-2**, the duplicated data is removed from the first storage device **210-1** and the first storage device **210-1** becomes associated with the second group of storage devices **220**.

[0037] The principles herein are implemented as hardware, firmware, software or any combination thereof. Moreover, the software is preferably implemented as an application program tangibly embodied on a program storage unit or computer readable medium. The application program may be uploaded to, and executed by, a machine including any suitable architecture. Preferably, the machine is implemented on a computer platform having hardware such as a processing unit ("CPU"), a memory, and input/output interfaces. The computer platform may also include an operating system and microinstruction code. The various processes and functions described herein may be either part of the microinstruction code or part of the application program, or any combination thereof, which may be executed by a CPU, whether or not such computer or processor is explicitly shown. In addition, various other peripheral units may be connected to the computer platform such as an additional data storage unit and a printing unit and/or display unit.

[0038] All examples and conditional language recited herein are intended for pedagogical purposes to aid the reader in understanding the principles of the disclosure and the concepts contributed by the inventor to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions. Moreover, all statements herein reciting principles, aspects, and embodiments, as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents as well as equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure.

What is claimed is:

1. A method for controlling a multi-tiered storage device (MTSD), the method comprising:

receiving a block of data using a controller of the MTSD, the MTSD comprising a first storage comprising one or more storage devices, and a second storage comprising one or more storage devices, wherein each storage device of the first storage has a higher write endurance than each storage device of the second storage;

writing, by the controller, the received block of data to a first storage device of the first storage;

assigning a timestamp respective of the received block of data; and

writing, by the controller, the received block of data to a second storage device of the second storage, in response to determining that a time elapsed since the timestamp is above a first threshold value.

2. The method of claim **1**, further comprising:

erasing the received block of data from the first storage device of the first storage, in response to determining that the received block of data was written to the second storage device of the second storage.

3. The method of claim **1**, wherein a storage device of the first storage is one from among a solid-state drive (SSD), and a hard-disk drive (HDD).

4. The method of claim **1**, further comprising:

detecting a degradation in a physical parameter of the first storage device of the first storage; and

associating the first storage device with the second storage.

5. The method of claim 4, further comprising:
duplicating at least a portion of data stored on the first storage device of the first storage to another storage device of the first storage; and
removing the duplicated data from the first storage device of the first storage.

6. The method of claim 4, wherein detection of the degradation in the physical parameter comprises:
determining that the physical parameter of the first storage device is below a second threshold value, the physical parameter comprises at least one of endurance, performance, update latency and access time.

7. A system for controlling a multi-tiered storage device (MTSD), the system comprising:
a processing unit (PU);
an input/output (I/O) interface communicatively coupled to the MTSD, the MTSD comprises a first storage comprising one or more storage devices, and a second storage comprising one or more storage devices, each storage device of the first storage has a higher write endurance than each storage device of the second storage; and
a memory communicatively coupled to the PU, the memory storing instructions that when executed by the PU configure the system to:
receive a block of data;
write the received block of data to a first storage device of the first storage;
assign a timestamp respective of the received block of data; and
write the received block of data to a second storage device of the second storage, upon determination that a time elapsed since the timestamp is above a first threshold value.

8. The system of claim 7, wherein the memory further stores instructions that when executed by the PU configure the system to erase the received block of data from the first storage device of the first storage, in response to determining that the received block of data was written to the second storage device of the second storage.

9. The system of claim 7, wherein the first storage device is one from among a solid-state drive (SSD), and a hard-disk drive (HDD).

10. The system of claim 7, wherein the memory further stores instructions that when executed by the PU configure the system to:
detect a degradation in a physical parameter of the first storage device of the first storage; and
associate the first storage device with the second storage.

11. The system of claim 10, wherein the memory further stores instructions that when executed by the PU configure the system to:
duplicate at least a portion of data stored on the first storage device of the first storage to another storage device of the first storage; and
remove the duplicated data from the first storage device.

12. The system of claim 10, wherein the memory further stores instructions that when executed by the PU to detect the degradation in the physical parameter further configure the system to determine that the physical parameter of the first storage device is below a second threshold value, the physical parameter includes at least one of endurance, performance, update latency and access time.

13. A non-transitory computer readable medium having stored thereon instructions for causing one or more processing units to execute the method according to claim 1.

14. A system comprising:
a controller; and
a memory that stores information arranging storage devices in a multi-tiered storage device (MTSD) into at least two groups including a first group of storage devices and a second group of storage devices, the storage devices in the first group of storage devices have a write endurance that is higher than a write endurance of the storage devices of the second group of storage devices,
wherein in response to the write endurance of a first storage device of the first group of storage devices falling below a threshold, the controller re-assigns the first storage device to the second group of storage devices.

15. The system according to claim 14, wherein in response to the write endurance of a first storage device of the first group of storage devices falling below a threshold, the controller further causes data on the first storage device to be written to a second storage device of the first group of storage devices.

16. The system according to claim 14, wherein the system is connected to the MTSD over a network.

* * * * *