METHOD FOR STORING DATA BY STORAGE DEVICE AND STORAGE DEVICE

TECHNICAL FIELD

[0001] The present invention relates to the field of information technologies, and in particular, to a method for storing data by a storage device and a storage device.

BACKGROUND

[0002] A solid state disk (Solid State Disk, SSD), as shown in FIG. 1, includes a storage controller 101 and a medium 102 (for example, a flash memory chip). The storage controller 101 includes a central processing unit ~~1011~~ (Centeral Processing Unit, CPU) 1011 and a memory 1012. Storage in the SSD is organized by using a physical block (physical block) and a page (page) as units. The page is the smallest read/write unit in the solid state disk, and a size of the page may be 4 KB, 8 KB, or 16 KB. Pages are combined into a physical block, and each physical block may have 32, 64, or 128 pages. The SSD generally divides storage space into data space and reserved space (Over-Provisioning). The data space is space to which data is already written, and the reserved space is free space, includes free pages, and data may be written to the reserved space. When data already stored in the data space in the SSD is to be overwritten by new data, a redirect-on-write (Redirection On Write, ROW) mechanism is used. That is, when the SSD writes new data to a logical block address (Logical Block Address, LBA) to modify the already stored data, the SSD writes the new data to a page of the reserved space, establishes a mapping relationship between the LBA and a page address of the reserved space, and marks data in a page address, to which the LBA is previously mapped, of the data space as garbage data. When the reserved space is less than a threshold, the SSD performs garbage space recycling for a physical block of the page in which the garbage data is located. A recycling process is as follows: reading valid data in the physical block of the page in which the garbage data is located, writing the read valid data to the reserved space, erasing data in the physical block of the page in which the garbage data is located, and using the physical block as new reserved space. In the garbage space recycling process, a process in which the valid data is read and the valid data is written to the reserved space is referred to as a movement of valid data.

[0003] The garbage space recycling causes write amplification, and a ratio of a sum of a size V

of the valid data moved in the garbage space recycling in the SSD and a size W of the newly written data to the size W of the newly written data, that is, (V+W)/W, is referred to as write amplification.

SUMMARY

[0004] According to a first aspect, an embodiment of the present invention provides a solution for storing data by a storage device, where the storage device includes a first storage area and a second storage area, where the first storage area includes data space and reserved space, and the second storage area includes data space and reserved space, the storage device receives a write request, where the write request carries a logical address and data, and the storage device determines a feature of the write request.

[0005] When the feature of the write request meets a first condition, the storage device writes the data carried in the write request to a first storage address of the reserved space of the first storage area, and establishes a mapping relationship between the logical address and the first storage address, or when the feature of the write request meets a second condition, the storage device writes the data carried in the write request to a second storage address of the reserved space of the second storage area, and establishes a mapping relationship between the logical address and the second storage address. In this embodiment of the present invention, the data carried in the write request is written to reserved space of different storage areas according to the feature of the write request, which reduces movements of valid data in a garbage space recycling process and reduces write amplification.

[0006] Optionally, the storage device may independently perform garbage space recycling for the first storage area and the second storage area. That the storage device may independently perform garbage space recycling for the first storage area and the second storage area means that the storage device performs garbage space recycling for one of the first storage area and the second storage area, and does not affect the other storage area, or may perform garbage space recycling for both the first storage area and the second storage area. For write requests having different features, data is written to different storage areas, and garbage space recycling is independently performed for the storage areas based on different reserved space configured in the different storage areas. Therefore, movements of valid data in a garbage space recycling process can be reduced, write amplification can be reduced, and a quantity of times of triggering the garbage space recycling process can also be reduced by configuring different reserved space, so that a quantity of times of erasing a physical block in the storage device is reduced, and a service life of the storage device is increased.

[0007] Optionally, a size of the reserved space of the first storage area is different from a size of the reserved space of the second storage area.

[0008] Optionally, the reserved space of the first storage area is smaller than the reserved space of the second storage area, or a ratio of the reserved space of the first storage area to the data space of the first storage area is less than a ratio of the reserved space of the second storage area to the data space of the second storage area. Because the second storage area has more reserved space, a quantity of times of garbage space recycling in the second storage area can be reduced.

[0009] Optionally, when data is written to a corresponding storage area, for example, the foregoing first or second storage area, if reserved space of the corresponding storage area is insufficient, dynamic adjustment may be performed. For example, under the precondition that a maximum redundant quota is not used up, extra reserved space may be applied for, and then the data in the write request is written to the reserved space that is newly applied for.

[0010] Optionally, the determining, by the storage device, a feature of the write request includes:

determining, by the storage device, whether the write request is a sequential write request or a random write request, where the first condition is the sequential write request, and the second condition is the random write request, and respectively storing the sequential write request and the random write request in different storage areas, which reduces movements of valid data in a garbage space recycling process and reduces write amplification. In addition, ~~in a case in which~~when both the sequential write request and the random write request exist, random write performance of the storage device is improved without affecting performance of the sequential write request. Optionally, it is determined whether data is written to a reference logical address, where an absolute value of an address difference between the reference logical address and the logical address is not greater than L, and L may be set according to a requirement for the sequential write request. In one implementation, when data is written to the reference logical address, the write request is a sequential write request; when no data is written to the reference logical address~~otherwise~~, the write request is a random write request. In another implementation, when no data is written to the reference logical address, the write request is a random write request. When data is written to the reference logical address, further, the storage device determines whether an interval between a time at which the write request carrying a reference address is received for the last time and a time at which the write request carrying the logical address is received for the last time is greater than a threshold T, and if when the interval is greater than the threshold T~~yes~~, the write request is still a random write request; ~~otherwise~~when the interval is not greater than the threshold T, the write request is a sequential write request. T may be set according to a specific implementation.

[0011] Optionally, the determining, by the storage device, a feature of the write request includes:

determining, by the storage device, a sequence level of the write request, where the first

condition is a first sequence level range, the second condition is a second sequence level range, and a minimum value of the first sequence level range is greater than a maximum value of the second sequence level range. Optionally, the sequence level is a ratio S (a sequence level) of a sequential write request count Cs of a current logical address to (a sum of the sequential write request count Cs of the current logical address and a random write request count Cr of the current logical address), that is, Cs/(Cs+Cr)=S. According to the sequence level, data carried in write requests having different sequence levels may be respectively stored in different storage areas having different reserved space, which reduces movements of valid data in a garbage space recycling process and reduces write amplification.

[0012] Optionally, the determining, by the storage device, a feature of the write request includes:

determining, by the storage device, a randomness level of the write request, where

the first condition is a first randomness level range, the second condition is a second randomness level range, and a maximum value of the first randomness level range is less than a minimum value of the second randomness level range. Optionally, the randomness level is a ratio R of a random write request count Cr of a current logical address to (a sum of a sequential write request count Cs of the current logical address and the random write request count Cr of the current logical address), that is, Cr/(Cr+Cs)=R. According to the randomness level, data carried in write requests having different randomness levels may be separately stored, which reduces movements of valid data in a garbage space recycling process and reduces write amplification.

[0013] Optionally, the determining, by the storage device, a feature of the write request includes:

determining, by the storage device, a size of the data carried in the write request, where

the first condition is a first data range stored in the first storage area, the second condition is a second data range stored in the second storage area, and a minimum value of the first data range is greater than a maximum value of the second data range. The storage device includes different storage areas, each storage area stores a corresponding data ~~storage~~ range, which may reduce movements of valid data in a garbage space recycling process and reduce write amplification. Optionally, a data range refers to an interval of a size of data carried in a write request stored in a storage area.

[0014] Optionally, in a case in which multiple write requests are concurrently sent, the storage device preferentially processes a write request meeting the first condition, to improve write performance of the storage device.

[0015] Optionally, the storage device is an SSD, or a shingled magnetic recording (Shingled

movement of valid data does not need to be performed, that is, there is no write amplification. Therefore, a sequential write request is stored in Vd1, and in the garbage space recycling process, movements of valid data and write amplification are reduced in Vd1. In addition, because there is no valid data or little valid data in the garbage space recycling process, a small amount of reserved space may be allocated to Vd1.

[0043] As shown in FIG. 6, in this embodiment of the present invention, Vd2 includes X physical blocks, where each physical block includes (n+1) pages, Vd2 is configured to store data carried in a random write request, a first physical block to an $(X-3)^{th}$ physical block form data space of Vd2, and an $(X-2)^{th}$ physical block to an X^{th} physical block form reserved space.

[0044] As shown in FIG. 7, an SSD receives a first random write request, where a logical address carried in the first random write request is an LBA 1'. The LBA 1' is already mapped to the first m pages of the first physical block in Vd2, that is, the LBA 1' is mapped to a page 0 to a page m–1 of the first physical block in Vd2. Based on a ROW mechanism of the SSD, the SSD writes data carried in the first random write request to the $(X-2)^{th}$ physical block of the reserved space in Vd2, establishes a mapping between the LBA 1' and the first m pages of the $(X-2)^{th}$ physical block in Vd2, that is, establishes a mapping between the LBA 1' and a page 0 to a page m–1 of the $(X-2)^{th}$ physical block in Vd2, and identifies data in the page 0 to the page m–1 of the first physical block in Vd2 as garbage data (and removes the mapping between the LBA 1' and the first m pages of the first physical block in Vd2). The SSD receives a second random write request, where a logical address carried in the second random write request is an LBA 3'. The LBA 3' is already mapped to the first (n+1–m) pages of the second physical block in Vd2, that is, the LBA 3' is mapped to a page 0 to a page n–m of the second physical block in Vd2. Based on the ROW mechanism of the SSD, the SSD writes data carried in the second random write request to the $(X-2)^{th}$ physical block of the reserved space in Vd2, establishes a mapping between the LBA 3' and the first (n+1–m) pages of the $(X-2)^{th}$ physical block in Vd2, that is, establishes a mapping between the LBA 3' and a page m to a page n of the $(X-2)^{th}$ physical block in Vd2, and identifies data in the page 0 to the page n–m of the second physical block in Vd2 as garbage data (and removes the mapping between the LBA 3' and the first (n+1–m) pages of the second physical block in Vd2).

[0045] As shown in FIG. 8, when the reserved space in Vd2 is less than a threshold and garbage space recycling is started, ~~it is determined that~~ physical blocks for which recycling needs to be performed are the first physical block and the second physical block in Vd2. Valid data is stored in a page m to a page n of the first physical block, and a movement of valid data needs to be performed. In this embodiment of the present invention, the valid data stored in the page m to the page n of the first physical block is moved to a page 0 to a page n–m of the $(X-1)^{th}$ physical block in Vd2, and a

includes R physical blocks, and each physical block includes (n+1) pages. In another implementation manner, each storage area may include a different quantity of physical blocks, which is not limited in this embodiment of the present invention. The SSD selects a storage area according to a size of data carried in a write request. Vd1 is configured to store (0-4 KB] data, Vd2 is configured to store (4 KB-8 KB] data, Vd3 is configured to store (8 KB-16 KB] data, Vd4 is configured to store (16 KB-32 KB] data, Vd5 is configured to store (32 KB-64 KB] data, Vd6 is configured to store (64 KB-128 KB] data, Vd7 is configured to store (128 KB-256 KB] data, Vd8 is configured to store (256 KB-512 KB] data, and Vd9 is configured to store data that is greater than 512 KB. The (4 KB-8 KB] data is also referred to as data of a data range. A data range refers to an interval of a size of data carried in a write request stored in a storage area. A data range of Vd1 represents that a size of data carried in a write request stored in Vd1 is not greater than 4KB. When receiving a write request, the SSD determines, according to a data range stored in each storage area, a storage area that is used to store data carried in the write request. For example, the second storage area shown in FIG. 2 is Vd1. Because data in the data range of Vd1 is not greater than 4 KB (the second condition shown in FIG. 2), data carried in multiple write requests is stored in a same physical block. When some of the data is identified as invalid data because of modification, and garbage space recycling is performed for a physical block, data carried in another write request stored in the physical block is used as valid data, and a movement of valid data needs to be performed, thereby causing write amplification. For example, Vd9 is the first storage area shown in FIG. 2. Because data in a data range of Vd9 is greater than 512 KB (the first condition shown in FIG. 2), data carried in a same write request or a small quantity of write requests is stored in a same physical block. Therefore, when the SSD receives again a write request for modifying data in the physical block, data stored in pages of the entire physical block is identified as garbage data. When garbage space recycling is performed for the physical block, because all or most of data in the entire physical block is garbage data, no movement of valid data is generated or only a small quantity of movements of valid data is generated, which does not cause write amplification or causes small write amplification. Therefore, the SSD is divided into different storage areas, and each storage area stores a corresponding data ~~storage~~ range, which may reduce movements of valid data in a garbage space recycling process, and reduce write amplification. Optionally, a size of data carried in a write request determines a corresponding storage area. More reserved space is allocated to a storage area corresponding to a data range storing small data than to a storage area corresponding to a data range storing big data, or a ratio of reserved space of a storage area corresponding to a data range storing small data to data space of the storage area corresponding to the data range storing small data is greater than a ratio of reserved space of a storage area corresponding to a data range storing big data

similar to that of a physical block in an SSD. Therefore, a solution in which movements of valid data are reduced ~~in~~ during garbage ~~data~~ space recycling in the SSD in the embodiments of the present invention may also be applied to the SMR disk. The SMR disk is divided into different storage areas, where each storage area includes multiple zones (including a data zone and a reserved zone), reserved zones having different sizes are allocated to the different storage areas, and a feature of a write request is determined. For example, whether the write request is a random write request or a sequential write request is determined, or a sequence level of the write request is determined, or a randomness level of the write request is determined, or a size of data carried in the write request is determined. The data carried in the write request is stored in a reserved zone of a specific storage area, so that movements of valid data in an SMR disk during garbage space recycling are reduced, and write amplification is reduced. For a specific implementation, reference may be made to an implementation solution of the SSD, and details are not described herein again in this embodiment of the present invention. For a manner in which reserved zones may be allocated to different storage areas in an SMR disk, refer to a manner described above in which reserved space is allocated to different storage areas in an SSD.

[0064] In addition, in a storage array having a garbage space recycling function and based on a ROW mechanism, for example, in an all-SSD storage array and a hard disk drive (HDD) storage array, a storage array controller divides a logical block address of each hard disk into blocks according to a unit (for example, 1 MB). One block is taken from each disk of N disks to form a segment (segment) meeting a condition (for example, a segment of a redundant array of independent disks (Redundant Array of Independent Disks, RAID), for example, to form a RAID 6 (including 3 data blocks + 2 check blocks). A sequential write manner is used in the segment to improve write performance. Data in the segment cannot be overwritten, and valid data in the segment needs to be first moved before writing. The storage array controller divides storage space into a data segment (a segment to which data is already written) and a reserved segment (a free segment), and records a mapping from a logical address to physical storage space. When data is being written, the storage array controller sequentially writes the data to a reserved segment, then maps a logical address to a physical address to which the data is newly written, and marks data stored in a physical address to which the logical address is previously mapped as garbage data. After a quantity of reserved segments is less than a threshold, garbage space recycling is started, a segment having the most garbage data is found, a valid data ~~block~~ in the segment is moved, and the segment becomes a reserved segment to which data may continue to be written.

[0065] In the foregoing storage array, the segment has a characteristic similar to that of a physical block in an SSD. Therefore, a solution in which movements of valid data are reduced

CLAIMS

What is claimed is:

1. A method for storing data by a storage device, wherein the storage device comprises a first storage area and a second storage area, wherein the first storage area comprises data space and reserved space, and the second storage area comprises data space and reserved space, and the method comprises:

receiving, by the storage device, a write request, wherein the write request carries a logical address and data;

determining, by the storage device, a feature of the write request;

writing, by the storage device when the feature of the write request meets a first condition, the data carried in the write request to a first storage address of the reserved space of the first storage area; and

establishing, by the storage device, a mapping relationship between the logical address and the first storage address;~~ or~~

writing, when the feature of the write request meets a second condition, the data carried in the write request to a second storage address of the reserved space of the second storage area; and

establishing, by the storage device, a mapping relationship between the logical address and the second storage address.

2. The method according to claim 1, wherein the reserved space of the first storage area is smaller than the reserved space of the second storage area.

3. The method according to claim 1, wherein a ratio of the reserved space of the first storage area to the data space of the first storage area is less than a ratio of the reserved space of the second storage area to the data space of the second storage area.

4. The method according to claim 2 or 3, wherein the determining, by the storage device, a feature of the write request comprises:

determining, by the storage device, whether the write request is a sequential write request or a random write request, wherein

the first condition is the sequential write request, and the second condition is the random write request.

5. The method according to claim 2 or 3, wherein the determining, by the storage device, a feature of the write request comprises:

determining, by the storage device, a sequence level of the write request, wherein

the first condition is a first sequence level range, the second condition is a second sequence