(12) UK Patent Application (19) GB (11) 2529669 (13) A

(43) Date of A Publication 02.03.2016

(21) Application No: 1415248.2

(22) Date of Filing: 28.08.2014

(71) Applicant(s):
International Business Machines Corporation
New Orchard Road, Armonk 10504, New York,
United States of America

(72) Inventor(s):
Giovanni Cherubini
Llias Lliadis
Jens Jelitto
Vinodh Venkatesan

(74) Agent and/or Address for Service:
IBM United Kingdom Limited
Intellectual Property Law, Hursley Park,
WINCHESTER, Hampshire, SO21 2JN,
United Kingdom

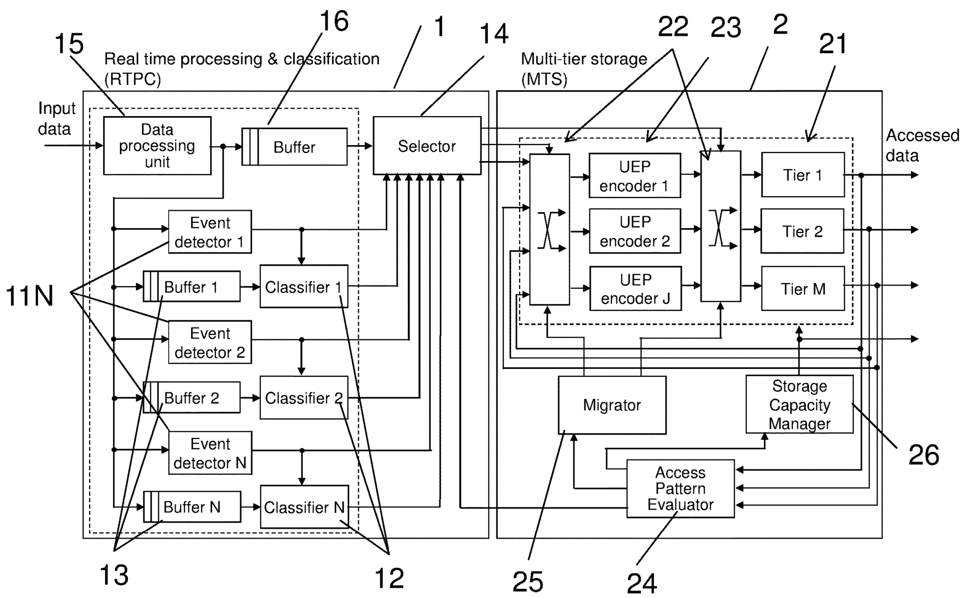
(51) INT CL:
G06F 3/06 (2006.01)

(56) Documents Cited:
US 8370597 B1 **US 8315995 B1**
US 20140281322 A1 **US 20120117328 A1**

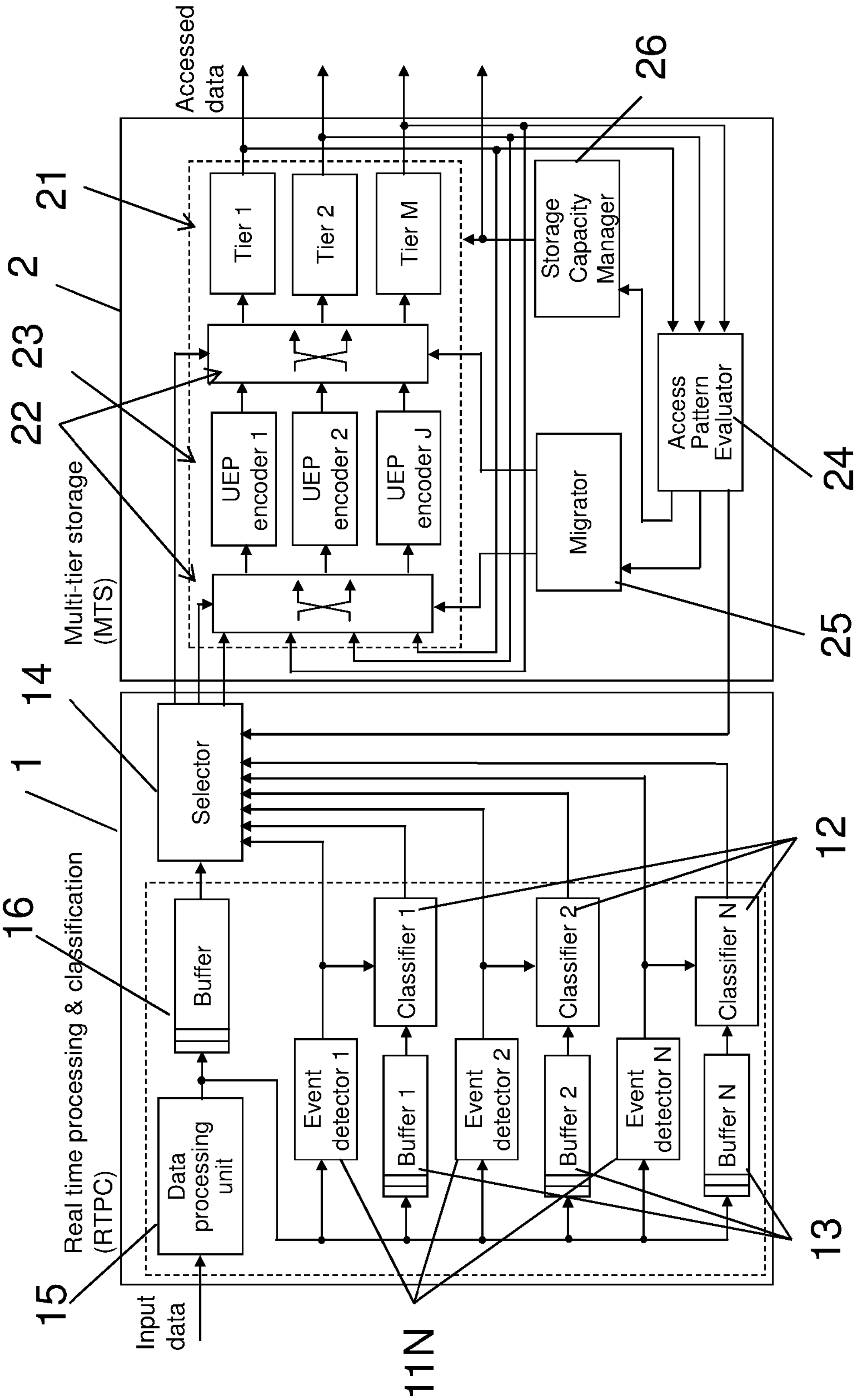
(58) Field of Search:
INT CL **G06F**
Other: **EPODOC, WPI, TXTE**

(54) Title of the Invention: **Storage system**
Abstract Title: **Multi-tiered storage system**

(57) A storage system comprises a storage unit 2 with at least two storage tiers 21, and an access pattern evaluator 24 configured to provide information about a frequency at which data segments stored in one or more of the at least two storage tiers are accessed. A classification unit 1 is configured to assign at least one out of a set of at least two relevance classes to a data segment received for storing in the storage unit dependent on information comprised in the data segment. A selector 14 is configured to determine a storage tier out of the at least two storage tiers for storing the classified data segment to dependent on at least access frequency information provided by the access pattern evaluator for data segments in the same relevance class, and configured to determine a level of protection for the classified data segment dependent on at least the relevance class assigned. Logic 22 is provided for storing the classified data segment including the assigned relevance class to the determined storage tier and according to the determined level of protection. The system can also have a storage tier relocation manager and storage capacity manager. The system is designed to deal with emerging big data applications.



GB 2529669 A


Accessed data
Multi-tier storage (MTS)
Tier 1
Tier 2
Tier M
UEP encoder 1
UEP encoder 2
UEP encoder J
Storage Capacity Manager
Migrator
Access Pattern Evaluator
Real time processing & classification (RTPC)
Selector
Buffer
Data processing unit
Event detector 1
Event detector 2
Event detector N
Buffer 1
Buffer 2
Buffer N
Classifier 1
Classifier 2
Classifier N
Input data
1
2
11N
12
13
14
15
16
21
22
23
24
25
26

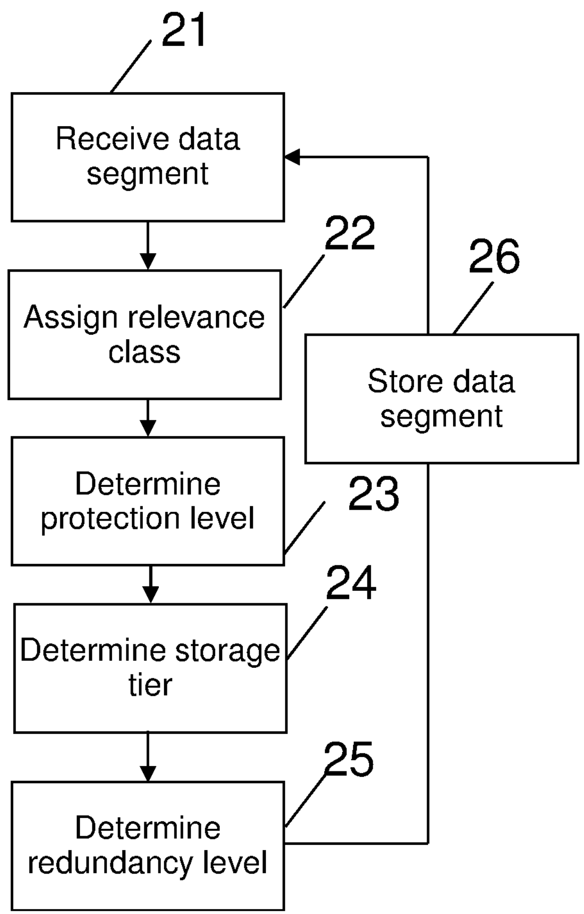


FIG. 2

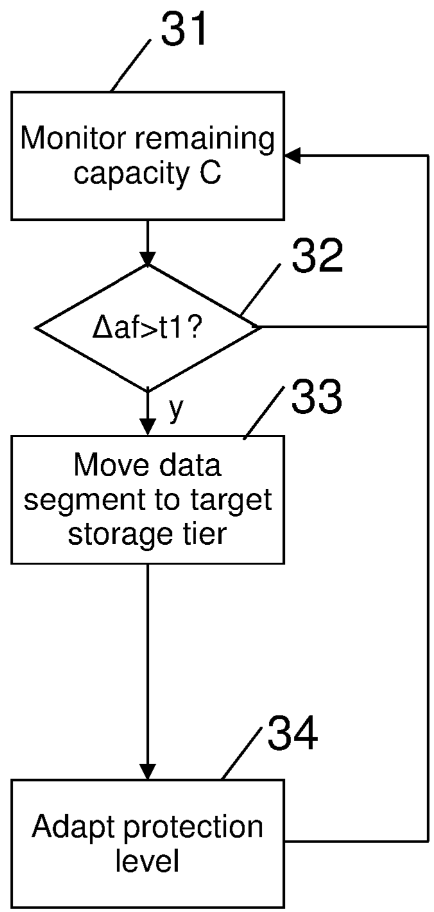


FIG. 3

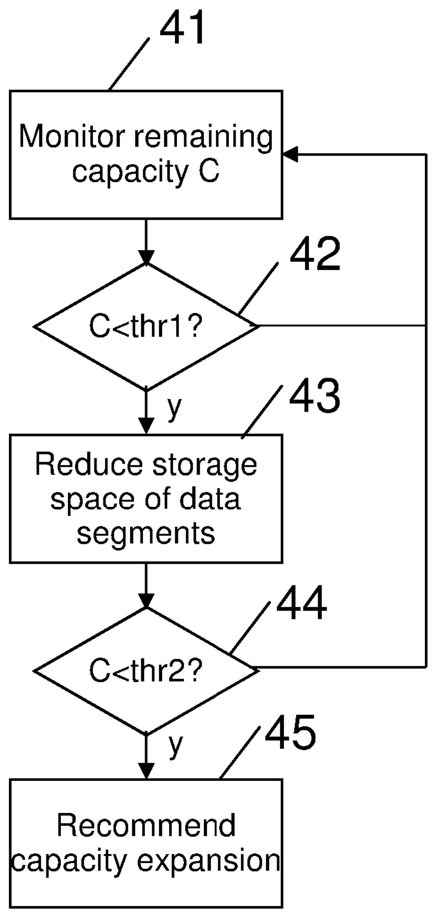


FIG. 4

STORAGE SYSTEM

FIELD OF THE INVENTION

The present invention relates to a storage system, a storage capacity manager, a storage tier relocation manager, a method for storing a data segment in a storage tier of a storage unit comprising at least two storage tiers, a method for managing a capacity of a storage unit for storing data segments, a method for relocating a data segment presently stored in a storage tier of a storage unit, and corresponding computer program products.

BACKGROUND

Today’s multi-tiered storage systems are especially suited for offering a trade-off between high performance and efficient low-cost long-term storage of data. However, very limited intelligence is usually available to determine without human intervention within which tier a certain data file should be stored. While today's approach may be adequate for most applications given the number and size of data files that need to be stored and retrieved, it appears that a new paradigm is needed to address the challenges posed by applications where a very large amount of data is to be stored and valuable information reliably is to be identified and accessed. Examples of so-called big data applications are emerging in various fields, including social networks, sensor networks, and huge archives of business, scientific and government records. One of the ultimate big data challenges, however, is represented by a Square Kilometer Array (SKA) telescope, expected to be completed in 2024, whose antennas will gather tens of exabytes of data and store petabytes of data every day. Another significant big data challenge lies in the healthcare industry, where personalized medicine and large-scale cohort studies may require storage of medical data for extended periods of time.

BRIEF SUMMARY OF THE INVENTION

According to an embodiment of one aspect of the invention, a storage system is provided comprising a storage unit comprising at least two storage tiers, and an access pattern evaluator configured to provide information about a frequency at which data segments stored in one or more of the at least two storage tiers are accessed. In a classification unit, at least one out of a set of at least two relevance classes is assigned to a data segment received for storing in the storage unit dependent on information comprised in the data segment. In a selector a storage tier out of the at least two storage tiers is determined for storing the classified data segment dependent on at least access frequency information provided by the access pattern evaluator for data segments in the same relevance class. In addition, a level of protection is determined in the selector for the classified data segment dependent on at least the relevance class assigned. Logic is provided for storing the classified data segment including the assigned relevance class to the determined storage tier and according to the determined level of protection.

In embodiments, the storage system may comprise one or more of the following features:

- the classification unit comprises a set of event detectors with each event detector of the set being configured to detect a different event in the data segment received for storing;
- the classification unit comprises a set of classifiers, each classifier being assigned to a different one of the event detectors, and each classifier of the set being configured to assign at least one out of one or more relevance classes pre-selected from the set of relevance classes for the event to be detected by the assigned event detector;
- a real-time data processing unit for real-time processing of input data segments and providing a sequence of data segments for storing;
- the selector is configured to determine the storage tier for the classified data segment dependent on the access frequency information provided by the access pattern evaluator for data segments in the same relevance class, and dependent on the relevance class assigned;
- wherein the selector is configured to determine the level of protection for the classified data segment dependent on the relevance class assigned and dependent on the access frequency information provided by the access pattern evaluator for data segments in the same relevance class;

- the selector is configured to determine a redundancy level for the classified data segment dependent on the determined protection level and the determined storage tier, the redundancy level specifying a number of copies of the classified data segment to be stored in the storage unit, and in particular specifying a number of copies of the classified data segment to be stored in which storage tier of the storage unit;
- the logic is configured to store copies of the classified data segment in the one or more storage tiers according to the determined level of redundancy;
- the selector is configured to determine one or more of an error correction code and an erasure code to be applied to the classified data segment dependent on the determined protection level and the determined storage tier;
- the logic is configured to store the classified data segment with the one or more of the determined error correction code and erasure code in the storage unit.

According to an embodiment of another aspect of the present invention, a method is provided for storing a data segment in a storage tier of a storage unit comprising at least two storage tiers. At least one out of a set of at least two relevance classes is assigned to the data segment dependent on information comprised in the data segment. Information is received about a frequency at which data segments stored in one or more of the at least two storage tiers are accessed. A storage tier out of the at least two storage tiers is determined for storing the classified data segment dependent on at least the access frequency information received for data segments in the same relevance class. A level of protection is determined for the classified data segment dependent on at least the relevance class assigned. The classified data segment including the assigned relevance class is stored to the determined storage tier and according to the determined level of protection.

According to an embodiment of a further aspect of the present invention, a storage tier relocation manager is provided for relocating a data segment presently stored in a storage tier of a storage unit with at least two storage tiers, and in particular stored in a storage system according to any one of the preceding embodiments, the data segment having assigned a protection level out of a set of protection levels. The storage tier relocation manager is configured to determine a target storage tier for the data segment dependent on access frequency information received for the data segment or for a relevance class to which the data segment is assigned. The storage tier relocation manager is further configured to relocate the data segment to the target storage tier if the target storage tier is different from the present storage tier, and in this case to apply a protection measure suitable for achieving the assigned protection level.

In a preferred embodiment, the storage tier relocation manager is configured to - in case of a relocation - to apply the protection measure by selecting one or more of: a redundancy level specifying a number of copies of the data segment to be stored in one or more of the target storage tier and a non-target storage tier; an error correction code to be applied to the data segment; and an erasure code to be applied to the data segment. The storage tier relocation manager is further configured to store the data segment in the storage unit according to the selected one or more of the redundancy level, the error correction code and the erasure code.

According to an embodiment of another aspect of the present invention, a method is provided for relocating a data segment presently stored in a storage tier of a storage unit with at least two storage tiers, and in particular stored in a storage system according to any one of the preceding embodiments, the data segment having assigned a protection level out of a set of protection levels. A target storage tier is determined for the data segment dependent on access frequency information received for the data segment or for a relevance class to which the data segment is assigned. The data segment is relocated to the target storage tier if the target storage tier is different from the present storage tier. In this case a protection measure is applied suitable for at least achieving the assigned protection level.

According to an embodiment of a further aspect of the present invention, a storage capacity manager is provided for a storage system comprising a storage unit for storing data segments having at least one relevance class assigned per data segment, and in particular for a storage system according to any one of the preceding embodiments. The storage capacity manager comprises a monitoring unit for determining if a utilization of the storage unit fulfils a criterion, and a capacity managing unit for, in response to the utilization of the storage unit fulfilling the criterion, selecting at least one data segment stored in the storage unit for one of a deletion thereof or a deletion of a copy thereof at least dependent on the relevance class assigned.

In embodiments, the storage capacity manager may comprise one or more of the following features:

- the storage capacity manager is configured to manage a storage unit comprising at least two storage tiers;
- the monitoring unit is configured to determine if the utilization of at least one of the at least two storage tiers fulfils the criterion;
- the capacity managing unit is configured to, in response to the utilization of at least one storage tier fulfilling the criterion, selecting the at least one data segment stored in this storage tier for one of a deletion thereof or a deletion of a copy thereof in the same storage tier or a different storage tier;
- the monitoring unit is configured to determine if the utilization of the storage unit falls below a capacity threshold;
- the capacity managing unit is configured to select, in response to the utilization of the storage unit falling below the capacity threshold, the at least one data segment;
- the capacity managing unit is configured to one of: suggest the at least one selected data segment for deletion or deletion of a copy thereof; suggest the at least one selected data segment for deletion or deletion of a copy thereof and delete the at least one selected data segment or a copy thereof in response to a user confirmation; delete the at least one selected segment or copies thereof;
- the capacity managing unit is configured to assign a retention class out of a set of retention classes to the data segments in the storage unit, each retention class out of the set indicating a measure for retaining the assigned data segment in the storage unit, the assignment of the retention class to a data segment being made dependent on the relevance class assigned to the data segment;
- the capacity managing unit is configured to select the at least one data segment for one of a deletion thereof or a deletion of a copy thereof dependent on the retention class assigned;
- the capacity managing unit is configured to select the at least one data segment for deletion if the corresponding data segment is within a number of n data segments showing the lowest retention classes assigned;
- the capacity managing unit is configured to manage the storage of data segments in a variable number of copies;
- the capacity managing unit is configured to select the at least one data segment for deleting one or more copies thereof if the corresponding retention class is below a retention threshold;

- the capacity managing unit is configured to determine the retention class to assign to a data segment in addition dependent on one or more of an age of the data segment; access frequency information for the data segment or for the relevance class the data segment is assigned to; a persistence index; a storage capacity available in one or more of the other storage tiers in case of a tiered storage unit.

According to an embodiment of another aspect of the present invention, a method is provided for managing a capacity of a storage unit for storing data segments having at least one relevance class out of a set of at least two relevance classes assigned per data segment, and in particular for a storage system according to any one of the preceding embodiments. It is determined if a utilization of the storage unit fulfils a criterion. In response to the utilization of the storage unit fulfilling the criterion, at least one data segment stored in the storage unit is selected for one of a deletion thereof or a deletion of a copy thereof at least dependent on the relevance class assigned.

According to an embodiment of a further aspect of the present invention, a computer program product is provided comprising a computer readable medium having computer readable program code embodied therewith, the computer readable program code comprising computer readable program code configured to perform a method according to any one of the preceding embodiments.

It is understood that method steps may be executed in a different order than listed in a method claim. Such different order shall also be included in the scope of such claim as is the order of steps as presently listed.

Embodiments described in relation to the aspect of an apparatus shall also be considered as embodiments disclosed in connection with any of the other categories such as the method, the computer program product, etc.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention and its embodiments will be more fully appreciated by reference to the following detailed description of presently preferred but nonetheless illustrative embodiments in accordance with the present invention when taken in conjunction with the accompanying drawings.

The figures are illustrating:

FIG. 1, a block diagram of a storage system according to an embodiment of the present invention,

FIG. 2, a flowchart of a method for storing a data segment in a storage tier of a storage unit comprising at least two storage tiers, according to an embodiment of the present invention,

FIG. 3, a flowchart of a method for relocating a data segment presently stored in a storage tier of a storage unit, according to an embodiment of the present invention, and

FIG. 4 a method for managing a capacity of a storage unit for storing data segments, according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE EMBODIMENTS

As an introduction to the following description, it is first pointed at general aspects of the invention.

<u>Storage tiers</u>

A storage system is understood as a tiered storage system once it comprises multiple tiers of storage. Different storage tiers preferably are embodied as storage devices of different technology, including but not limited to tape storage technology, hard disk drive (HDD) storage technology, solid state drive (SDD) storage technology such as flash storage, etc. The storage devices may offer different characteristics per storage tier, that may e.g. include storage volume, reliability e.g. measured in form of bit error rates, performance including access time, cost, term of storage, etc., such that when combining different storage technologies into a tiered storage system, considerable advantages can be achieved given that storage devices with different characteristics can be selected for storing data segments subject to the needs of the different data segments to be stored. Hence, it is preferred that in a multi-tiered storage system each storage tier comprises only one type of storage device. However, in a different embodiment different tiers of a tiered storage may even be based on the same storage technology, however, may still show different characteristics owed to the usage of storage devices of different generations, of electrical connections of different quality, etc. such that these storage devices differ in at least one characteristic that may be relevant for storing data segments and as such may impact the decision on which kind of storage device to place a data segment.

Generally, the different storage tiers may not necessarily reside at a common location or in a common housing but may be distributed as long as the classification unit to be introduced later on has access to the storage tiers and can store data segments to and retrieve data segments from the various storage tiers. Each storage tier as such may contain one or more physical devices. For example, an HDD storage tier may contain up to hundreds HDD, or in a different embodiment, only a single HDD.

Data segments

Data segments shall include any unit of data to be stored and in case of a tiered storage system any unit of data that may be a subject of an individual decision as to where in the tiered storage it is desired to be stored. Data segments may include one or more of blocks, pages, segments, files, objects files, portions of a data stream, etc.

Access pattern evaluator

An access pattern evaluator is configured to monitor accesses to the data segments stored in the storage unit, i.e. in at least one and preferably all of the various storage tiers. Accesses may in particular encompass read and / or write operations on a data segment, e.g. when a user of the storage system reads data segments from the tiered storage. The access pattern evaluator is configured to output such access patterns in form of access frequencies for the data segments. In a preferred embodiment, the access frequencies are not individually monitored and supplied, but access frequencies are provided with respect to relevance classes into which the stored data segments are classified as will be explained later on. The access pattern evaluator may be embodied in one of hardware, software or a combination thereof.

It is noted that the access pattern evaluator provides statistical data in form of an access frequency, for example. However, any other statistical data referring to accesses of stored data segments is meant to be subsumed under the term access frequency. The access pattern evaluator provides access frequencies for stored data segments, and as such evaluates a popularity of the data segments since the access frequency may be regarded as a measure of the popularity of a data segment or a relevance class into which the data segment is classified. The higher the access frequency the more popular the data segment or the corresponding relevance class is.

Classification unit

A classification unit as used in the context of the present invention is configured to classify a data segment that is requested to be stored in the storage unit into a relevance class. In one embodiment, two or more relevance classes may be assigned to a data segment, although the description mostly refers to one relevance class being assigned. A set of relevance classes from which a relevance class is selected for assignment preferably comprises at least two relevance classes. The classification of a data segment into a relevance class preferably is based on information comprised in the data segment to be classified; this information is also referred to as the content of the data segment. Hence, it is the content of the data segment that is evaluated for performing the classification. However, in another embodiment, information comprised in other data segments - e.g. data segments that are linked in time or space to the data segment to be classified - may also be evaluated for assigning a relevance class to the data segment.

The classification unit may take different embodiments. Subject to the complexity of information in the data segments and the number of data segments arriving for storage, it may be preferred, that an event detector is provided. Such an event detector may evaluate the data segments to be classified for an occurrence of one or more pre-defined events. An event detector may evaluate a data segment on its own, or multiple data segments in combination. In an example of data segments representing images supplied by a telescope, an event may be considered as the occurrence of an astronomical event such as the occurrence of a planet in the image. A classifier may then classify the event in more detail, such as in one or more of size, shape, color, etc. In this respect, the event detector may also be understood as a pre-classifier which limits the number of relevance classes available for this particular event down to a subset. The subsequent classifier then may only assign one or more relevance classes of this subset.

In another embodiment, multiple event detectors are provided and specifically each event detector is configured to detect a specific event that is different from the events the other event detectors are configured to detect. This arrangement is preferred in case parallel processing is required for big data applications. In such an embodiment, the classifier may be responsible for further classifying the detected events. However, it is preferred that one classifier is assigned to each event detector such that the number of classifiers corresponds to the number of event detectors. Hence, the classification may also be also parallelized. In yet another embodiment, multiple classifiers may be provided in combination with only a single event detector configured to detect different multiple events. In this case, the number of classifiers may correspond to the number of events that can be detected by the single event detector.

In another embodiment, the two step event detection and classification process may be replaced by a single classification step in which the data segments, which are input to the storage system, are evaluated versus the complete set of relevance classes. In the exemplary astronomical application, rather than looking for an astronomical event first and then classifying this event in more detail, the classification may be applied without prior event detection. Either way may result in the very same assignment of relevance class/es, e.g. “planet of size x and color y”. In a different embodiment, the classification unit may solely comprise an event detector which at the same time acts as a classifier specifically when one or more dedicated classes are assigned to event a priori. In a different view, the event may be known a priori, and only event features are to be identified, in which case an event detector is not needed.

The set of relevance classes available for tagging a data segment may be defined up-front and may be fixed and limited in size, or may change dynamically during operation of the storage system and/or during operation of a user application making use of the storage system. In one embodiment, self-learning algorithms are applied for changing and/or refining the set of relevance classes. For the overall storage system, it is envisaged that at least two relevance classes are provided and available for tagging the data segments. Subject to the complexity of the user application, hundreds of relevance classes may be available. In the case of a use of one or more event detectors, the corresponding subsets of relevance classes that are assignable for a particular event are defined up-front. A subset may at minimum contain one relevance class in case the event is sufficiently defined by such relevance class.

Each relevance class may imply a certain relevance of the data segments being classified thereto, wherein some relevance classes may refer to data segments with a content considered more important than the content of data segments assigned to other relevance classes. However, in a preferred embodiment, a relevance class may in the first instance solely represent a description of the content of the subject data segment, such as in the above example “planet of size x and color y”. Here, the relevance class assigned can rather be regarded as a descriptor for the content of the subject data segment. A relevance for the content described by the descriptor may at least later be added, e.g. by ranking the descriptors in order of relevance for storage purposes. Hence, the assignment of a relevance class to a data segment may in one embodiment include a mapping of descriptors to relevance classes, and e.g. also include a mapping of multiple different descriptors to a common relevance class. Finally, it is desired that a classification is applied that at least to some extent assigns a metric to a data segment reflecting an importance of the content of the data segment for storage purposes.

The classification unit assigns one or more relevance classes to a data segment to be stored, which is considered to be equivalent to assigning the data segment to one or more relevance classes. Preferably, only one relevance class is assigned per data segment. It is preferred that all data segments requested to be stored are classified and labelled by at least one relevance class. However, there may be envisaged a processing unit that preprocesses data segments arriving at the storage system. Such preprocessing may in one embodiment already lead to a selection of data segments to be stored out of all arriving data segments. In one embodiment, the processing unit is a real-time data processing unit for real-time processing of arriving data segments, also referred to as input data segments, e.g. in form of an input data stream, where the real-time data processing unit supplies a sequence of data segments to be stored, which are subsequently classified. Specifically, such a processing unit may apply one or more of filtering operations, suppression of spurious data segments, removing interference data segments, etc.

Under the assumption that the data segments are provided to the storage system as an input data stream, it is preferred that one or more buffers be provided in the classification unit in order to temporarily buffer the incoming and/or pre-processed data segments for providing sufficient time for conducting the classification and the determination of protection level and storage tier as will be explained later on. Hence, in one embodiment, a buffer is provided for buffering a data segment received for storage for at least a period of time required by the classification unit for assigning a relevance class to this data segment. In case the classification is implemented by multiple classifiers, one buffer may be provided per classifier, or a common buffer may be provided for more or all classifiers. In addition, or independent from the above buffer/s, another buffer preferably is provided for buffering a data segment received for storage for at least a period of time required by the selector for determining the storage tier and the protection level. After the determination, the selector may forward the data segment to be stored together with the class information and information as to the determined protection level and information as to the suggested storage tier to the storage unit, and in particular its logic.

Selector

The storage system further comprises a selector. The selector may be implemented in hardware, software or a combination of both. For every data segment to decide on, the selector receives, possibly amongst others, the relevance class assigned to the data segment by the classification unit and the access frequency information provided by the access pattern evaluator for data segments, and preferably for the particular relevance class to which the present data segment is assigned. Based on this information, i.e. the relevance class and the access frequency information for this relevance class, the selector determines a level of protection the data segment is to be stored with, and a storage tier in which the data segment is to be stored.

Hence, the selector takes its decision at least based on the content of the data segment, whose content is mapped into a relevance class as explained above. In one embodiment, a data segment with a high-rank relevance class is associated with a high value. Accordingly, its content is such that a loss would be associated with a high cost. Therefore, this data segment deserves a higher level of protection compared to data segments with lower value.

In a preferred embodiment, protection of a data segment in the present context may be oriented along the following types of impairment categories that a data segment may incur:

(a) data corruption where bits are altered,

(b) data erasure where bits are lost, and

(c) temporary data unavailability.

The corresponding metrics for these types of impairments may include:

- for type (a) impairments, a bit error rate metric;
- for type (b) impairments, a mean time to data loss (MTTDL) metric or a mean annual amount of data lost (MAADL) metric;
- for type (c) impairments, a percentage of time a data segment being unavailable;

These metrics again can be implemented by protection measures including one or more of the following:

- for type (a) impairments, a required bit error rate metric can be achieved by applying an error correction code of a given correction power to the data segment;
- for type (b) impairments, a mean time to data loss (MTTDL) metric or a mean annual amount of data lost (MAADL) metric can be achieved by applying an erasure code of a given correction power to the data segment;
- for type (c) impairments, a percentage of time a data segment being unavailable can be limited by providing copies of the data segment in the storage unit, also referred to as applying a redundancy level.

Hence, the protection level to be assigned may be selected from a set of protection levels available. In a preferred embodiment, each protection level is defined by a combination of individual impairment levels not to be underrun in the various impairment categories. The protection level then is achieved by a protection measure that addresses the one or more individually allowed impairment levels by corresponding one or more measures or a combination thereof, or by selecting a suitable redundancy level for the data segment; selecting a suitable error correction code for the data segment; selecting a suitable erasure code for the data segment. The determination of the redundancy level may in one embodiment specify the number of copies of the data segment is to be stored in one or more of the storage tiers.

Through the monitoring of data segment accesses by the access pattern evaluator it can be determined which relevance classes or which corresponding data segments are more popular than others. Every time a data segment is accessed, the associated metadata information preferably including the relevance class is provided to the access pattern evaluator, which learns about a popularity of the information content in the data segments from the way they are being accessed. Access patterns may be found at various levels including one or more of activity during various times of a day, sequence of reads and writes, access sequentiality, or number of users retrieving the data. This information is used to preferably further classify data segments into one of several popularity classes and shall also be subsumed under the access frequency information.

Any time the access frequency information changes, e.g. the popularity class changes, such a change may be sent by the access pattern evaluator to the selector, which accordingly may update a metric for an initial decision on a level of protection and a storage tier of individual input data segments. Therefore, the selector determines the data segment placement in the tiered storage and the level of protection based on both a data relevance classification and data access statistics. In this manner, a data segment that belongs to a certain relevance class is passed out to a suitable storage tier and is protected by means at least achieving the required protection level, e.g. by one or more of applying an error correction code, applying an erasure code, and a redundancy level that are most appropriate at a particular point in time.

In an embodiment, the selector assigns a protection level and a storage tier placement to an incoming data segment d_n at time nT, where 1/T is the rate at which data segments are received, based on metrics that depend on two variables, named "relevance index" $i_r(c_k)$, and "popularity index" $i_{p,n}(c_k)$, where c_k indicates the relevance class such that $d_n \in c_k$. Both $i_r(c_k)$ and $i_{p,n}(c_k)$ are real valued in the interval [0, 1]. Hence, it is apparent that classes are not necessarily restricted to discrete levels but also may be represented by real values as allowed. Note that the cardinality of the set of relevance classes is equal to $\aleph$, given by $\aleph=(K_1+1)x(K_2+1)x...x\ (K_N+1)$, where K_ℓ denotes the number of classes of the ℓ-th classifier, $\ell=1,2,...,N$. The relevance index corresponds to the importance of the relevance class as identified by the N classifiers, whereas the popularity index corresponds to the popularity of the relevance class as determined by the access pattern evaluator. The popularity index of each class varies over time depending on the access pattern, whereas the relevance index varies slowly compared to the popularity index, as a result of a varying assessment of the relevance of a class. It is assumed that at each time interval a new data segment is received, sufficient capacity is available at each storage tier for a new data segment allocation. The relevance class of a new data segment at the n-th time interval, denoted by d_n, is assigned by a classifier, or by the classification unit as such, and the relevance index is given by $i_r(c_k)$, where $d_n \in c_k$. As the data segment d_n is new to the system, its popularity class is ideally chosen as the most likely popularity class given that it belongs to relevance class c_k or may be assigned manually by an administrator or user.

An estimate of the most likely popularity index for a data segment that belongs to a certain relevance class may be obtained by updating at each time interval the popularity index estimate for each relevance class as

$i_{p,n}(c_k) = \max(i_{p,n-1}(c_k) - \epsilon_0, 0)$, if no data segment of class c_k is retrieved at the (n-1)-th time interval, or

$i_{p,n}(c_k) = \min(1, i_{p,n-1}(c_k) + \epsilon_1)$, otherwise,

where ϵ_0 and ϵ_1 are constant parameters.

In the absence of data access statistics, e.g. at initialization of a storage system, a correspondence between classes and storage tiers may be initially assumed, i.e. the higher the relevance class, the higher the storage tier wherein a hierarchy of the storage tiers is applied according to a single one or a combination of characteristics of the different storage tiers. E.g., a storage tier may be higher in the tier hierarchy if it provides faster access times, etc. However, some time after initialization of the storage system, additional information about the popularity of the data segments associated with a certain class is generated due to data retrieval activity and this may impact the selection of the storage tier, e.g. the higher the popularity of a relevance class, the higher the storage tier to which a new data segment in this class is assigned. Again, the storage tier may be regarded as superior in the tier hierarchy if it provides faster access times, for example.

In a preferred embodiment, an assignment of a tier placement $T(d_n)$ follows:

$T(d_n) = f_t(i_r(c_k), i_{p,n}(c_k))$,

and specifically

$$T(d_n) = f_t(\rho\, i_r(c_k) + \sigma\, i_{p,n}(c_k)) \quad (1)$$

and an assignment of a protection level $Q(d_n)$ follows:

$$Q(d_n) = f_q(i_r(c_k)\, ,\, i_{p.n}(c_k)) \quad (2)$$

In a preferred embodiment, a redundancy level $U(d_n)$ is assigned to a data segment d_n as follows:

$$U(d_n) = f_u(Q(d_n), T(d_n)) \quad (3)$$

where f_t and f_q are functions that univocally map a metric value to a tier level and to a protection level, respectively, ρ and σ are given system parameters, and f_u is a function that maps a tier and a protection level to a redundancy level. Hence, the determination of both a storage tier and a protection level for a data segment is preferably dependent on both the relevance index and the popularity index.

	Associated with	**Determined by**
Protection Level	Data segment	Relevance (or importance) of data segment
Reliability Level	Tier	Failure characteristics of the devices in the tier
Redundancy Level	Data segment	Protection level of the data segment and the reliability level of the tier(s) in which it is stored

Table I

Table I illustrates of the dependencies of the various levels. While in this embodiment, the protection level is solely dependent on the relevance class, the redundancy level, as the sole or one of more protection measures for implementing the assigned protection level, is dependent on this very protection level assigned as well as on the determined storage tier. For quantifying the redundancy level, the selected storage tier is preferably represented by its reliability which may be classified into a reliability level out of a set of reliability levels given that each type of storage device differs in particular in reliability, e.g. expressed by a bit-error rate. E.g., the bit error rate of tape is currently in the order of 1e-19, whereas that of HDDs is in the order of 1e-15.

Preferably, it is assumed that any storage tier selection is inherently dependent not only on the parameters assigned to the data segment to be stored, but also on the specifics of the storage tier, which in one embodiment may be represented by the reliability level into which its bit error rate may be classified.

The above equations (1) and (2) preferably implement one or more of the following characteristics:

a) The more relevant (or important) a data segment is, e.g. expressed by its associated relevance index, the higher its assigned level of protection;
b) The more popular (or frequently accessed) a data segment is, e.g. expressed by its associated popularity index, the faster the access it requires, i.e. the faster the storage tier that is selected for storage.

The following Table II illustrates an assignment of a storage tier and a protection level to a data segment according to this embodiment of the present invention:

	Relevant data	**Less important data**
Popular data	High level of protection Faster tier	Low level of protection Faster tier

Infrequently accessed data	High level of protection Slower tier	Low level of protection Slower tier

Table II

A protection level may be implemented by applying one or more of defined error correction code/s to the data segment, applying an erasure code across devices – such as RAID for HDDs - or storing the data segment a number of times in the same or in different tiers for providing redundancy. A combination of the means applied is also referred to as protection scheme or protection measure.

The following Table III illustrates an assignment of a storage tier and a protection level according to an embodiment of the present invention, wherein the protection level of a data segment is determined by the relevance class assigned, and wherein the storage tier is selected e.g. dependent on the access frequency information for the subject relevance class. Hence, the less frequently the data segment is accessed the lower is the storage tier in which the data segment is stored. However, a lower storage tier may not only be slower in access time but also be less reliable. Or, the data segment may a priori be assigned to a less reliable storage tier in view of more preferred storage tiers already being occupied. Still, the requested protection level may still be achieved via determining a suitable redundancy level. According to Table III, relevant data segments that require a high level of protection may therefore be stored on a less reliable storage tier, however, in multiple copies in this storage tier thereby providing a high level of redundancy. Alternatively, the data segments requiring a high level of protection may be stored in a more reliable storage tier, however requiring only a moderate number of copies in this storage tier, i.e. a moderate level of redundancy. In a third alternative, multiple copies of such data segments may be stored across multiple tiers.

	Relevant data	Less important data
Less reliable tier	High level of protection High level of redundancy	Low level of protection Moderate level of redundancy
More reliable tier	High level of protection Moderate level of redundancy	Low level of protection Low level of redundancy
Multiple tiers	High level of protection Multiple copies stored across tiers	Low level of protection Fewer copies stored across tiers

Table III

After the various determinations, the selector may forward the data segment to be stored together with the relevance class information, the protection level information and information as to the suggested storage tier to the storage unit, and in particular its logic. In a preferred embodiment, the required protection measure is also already determined by the selector and submitted to the logic.

<u>Logic</u>

Logic is provided for storing the data segment in the determined storage tier and for implementing the determined level of protection. The protection level may therefore in one embodiment be translated into a protection measure including one or more of storing a number of copies, also referred to as redundancy level, selecting an error correction code, or selecting an erasure code. Alternatively, if the protection measures are already determined by the selector, the logic may apply these protection measures. The logic may be implemented in hardware, software or a combination of both and is meant to be the entity executing the suggestion taken by the selector.

Stored data segment

A data segment finally stored in the assigned storage tier is preferably stored together with the assigned relevance class and the assigned protection level. These levels may be stored in combination with other metadata for the specific data segment.

Storage relocation manager

According to a preferred embodiment of the storage system, however, also as an aspect independent from the previously introduced embodiments of the storage system and the corresponding storage unit, a storage relocation manager is introduced since in a dynamic storage system, the popularity of each data segment as well as its relevance - although to a lesser extent - may change over time. Hence, a unit referred to as storage relocation manager may be in charge for moving data segments to other storage tiers of the storage unit, also referred to as target storage tiers. For example, when the popularity of a data segment increases it may be desirable to move it from a present slow storage tier to a faster storage tier to enable quicker access. When the popularity of a data segment decreases it may be desirable to move it from a fast present storage tier to a slower storage tier to free up space for other popular data segments. However, any movement solely based on the popularity index may have impact on the protection level, too, e.g. when the target storage tier has a different reliability than the present storage tier. The same is true when a data segment is replicated across multiple tiers.

In a preferred embodiment, the storage relocation unit, which is also referred to as the migrator, receives information from an access pattern evaluator such as described above and as such receives access frequency information for the individual relevance classes. This access frequency information enables the migrator to place data segments in the right storage tier to enhance access performance. Specifically, the migrator may move data segments stored in a present storage tier to another storage tier if such movement is indicated by the present access patterns of such data segments, and specifically by the access patterns of the class in which the respective data segment belongs. In another embodiment, the migrator may in addition monitor a relevance class assigned to the data segment and, specifically a change in such relevance class, which may also lead to a relocation of the data segment to a different storage tier.

It may be desirable that more relevant and popular data segments deserve a higher level of protection. To ensure a certain protection level in a given storage tier, a protection scheme is employed which is understood as a combination of protection measures to implement the desired protection level. The protection scheme may entail a combination of error correction codes within devices - e.g. for type (a) impairments -, erasure codes across devices - e.g. for type (b) impairments -, and replication across devices - e.g. for type (c) impairments as laid out above. However, when observing a different access frequency than in the past which may advise to move a data segment to a different storage tier, i.e. the target storage tier, the protection level in the target storage tier may be different than that of the present storage tier. If, on the other hand, the relevance of the data segment has not changed, the protection scheme preferably is to be amended. This is already because different storage tiers exhibit different levels of reliability, e.g., the bit-error rate of tape is $1e^{-19}$ whereas that of HDDs is $1e^{-15}$. Consequently, when moving data segments from one storage tier to another, the migrator preferably adapts the applied protection scheme in order to maintain the same protection level, e.g., by one or more of changing between 2-way versus 3-way replication, applying error correction and/or erasure codes with different number of parities, etc.

For each data segment d_ℓ, $\ell = 1, \ldots, L$, stored in the storage unit, an access pattern evaluator such as the one described above preferably assigns a popularity class c'_j and an associated popularity index $i_p(c'_j)$ which are determined by the number of accesses and the amounts of data read and written to each data segment in the recent history of time period T_1. The popularity class of each data segment is periodically sent by the access pattern evaluator to the migrator with time period T_2. The migrator then uses this information along with the relevance index c_k of each of each segment to determine a target tier $T_n(d_\ell)$, the new protection level $Q_n(d_\ell)$, and the new redundancy level $U_n(d_\ell)$ for that data segment for the time period nT_2 to $(n+1)T_2$ using expressions similar to (1), (2), and (3), e.g.:

$$T_n(d_\ell) = f_t(i_r(c_k), i_p(c'_j), C_1, \ldots, C_M, P_1, \ldots, P_M), \quad (4)$$

$$Q_n(d_\ell) = f_q(i_r(c_k), i_p(c'_j)), \quad (5)$$

$$U_n(d_\ell) = f_u(Q_n(d_\ell), T_n(d_\ell), R_1, \ldots, R_M), \quad (6)$$

wherein

- f_t and f_q are functions e.g., linear, that univocally map a metric value to a storage tier and to a protection level, respectively, and
- f_u is a function that maps a tier and a protection level to a redundancy level.

Here, $C_1, \ldots, C_M$ are the costs per gigabyte, $P_1, \ldots, P_M$ are the power consumption of a device, and $R_1, \ldots, R_M$ are the reliability indices which are metrics for the levels of reliability for each of the M tiers.

In one embodiment, for cost reduction, a certain protection level may be guaranteed by placing copies of data segments across multiple storage tiers. For example, a data segment with high relevance index and low to moderate popularity index may have one replica on an HDD storage tier for performance purposes, and another replica on a tape storage tier for reliability and cost purposes. It is known that erasure codes may provide much higher storage efficiency than replication for the same level of reliability. On the other hand, erasure codes may suffer from reduced access performance. Therefore, depending on the relevance and popularity indices, a choice may be made between an erasure code and replication based on the trade-off between storage efficiency and performance.

As described in connection with the selector, it is preferred that the Tables I, II and III also apply in the migration of already stored data segments, preferably in connection with the level of protection being specified for data segments in terms of certain metrics, e.g., MTTDL, availability, delay, etc, which can be associated with relevance classes and preferably popularity; in connection with the level of device reliability being specified in some metric (MTTF, ...) such as a failure or error characteristics of the storage devices/tiers, and in connection with the level of redundancy specifying parameters of an underlying redundancy scheme. The levels of protection for the data segments and the levels of device reliability for the device/s used within each tier are preferably known prior to a data segment replacement. The levels of redundancy are preferably determined such that the protection level for each data segment is guaranteed when the data segment is placed in a target tier.

The process introduced above is also referred to as dynamic tiering and may typically occur over large time scales compared to a time interval T over which a data segment is received for storage. The policies, according to which data is moved across different storage tiers and hence different types of storage devices, depend on access pattern characteristics and in addition preferably on the assigned relevance class. Depending on the storage device performance characteristics, certain tiering strategies may be better for a given workload than others. For instance, data segments accessed sequentially are preferably placed on HDDs, whereas randomly accessed data are preferably be placed on SSDs. Also, it is conceivable that the updated information regarding the popularity of the data segments associated with the various relevance classes in one embodiment is used to determine subsequent data segment movements. This, in turn, can steer the employment of effective caching and tiering strategies that have a significant effect on the cost and performance of the storage system.

Storage Capacity Manager

According to a preferred embodiment of the storage system, however, also as an aspect independent from the previously introduced embodiments of the storage system and the corresponding storage unit, a storage capacity manager is introduced because of the finite capacity of the storage unit, and a foreseen large amount of data segments steadily created within a big data system, which will likely make it necessary to discard obsolete data segments and/or to judiciously increase the storage system capacity of the storage unit. The storage capacity manager preferably has the main functionality of avoiding a storage unit capacity overflow by suggesting deleting the least relevant data segments from the storage unit, and / or by reducing a redundancy of data segments, i.e. deleting one or more copies of one or more data segments, and in particular by deleting one or more copies of one or more data segments belonging to a certain relevance class, and / or by providing recommendations to a system administrator for a capacity extension of the storage unit. For instance, whenever the stored data segments approach an available capacity of the storage unit which may be considered as a criterion of a utilization of the storage unit being fulfilled for initiating action the fulfilling of which criterion is monitored by a monitoring unit, and in particular if new storage capacity cannot be made available, a capacity managing unit of the storage capacity manager may select one or more data segments stored in the storage unit and may suggest these data segments or copies thereof for removal, i.e. erasure from the storage unit, or delete the selected one or more data segments or copies thereof, or delete the selected one or more data segments or copies thereto after having suggested for deletion to a user or to an administrator and after having received a confirmation for doing so.

The storage capacity manager may act on an individual storage unit such as an HDD, a tape, or an SDD, and as such detached from the previously described multi-tiered storage unit. However, in case of the storage unit comprising multiple storage tiers, the storage capacity manager may act on each storage tier individually or on the storage unit as a whole. Hence, the utilization of the storage unit fulfilling a criterion such as falling below a capacity threshold and therefore indicating a shortage of storage capacity in the storage unit, may refer to an individual tier of the storage unit or to the overall storage unit. Hence, in one embodiment, it may suffice that the monitoring unit detects an individual storage tier falling short of free capacity and therefore triggering a selection process for finally suggesting and / or deleting selected data segments in this specific storage tier. In another embodiment, the criterion may be set such that the total capacity of the storage unit including the multiple storage tiers is compared to a capacity threshold and initiates the selection process. In yet another embodiment, the detection of the storage capacity of an individual storage tier falling below a capacity threshold may trigger a selection of data segments out of the entire storage unit not limited to the data segments stored in the storage tier that falls short in free capacity. It is noted that in the case of a tiered storage unit thresholds indicating a shortage of free capacity may be set different for different storage tiers.

The monitoring unit for monitoring the fulfillment of the criterion related to the storage capacity of the or of a part of the storage unit may be embodied as hardware or software or a combination of both. The utilization of the storage unit may in one embodiment be represented by the still available storage capacity of the respective unit or of an individual storage tier, or by the utilized, i.e. occupied and / or reserved storage capacity of the storage unit or by an individual storage tier. Preferably, the criterion indicates a shortage of still available storage capacity in the respective storage tier or unit. In another embodiment, the criterion may be a rate at which new data segments are stored in the storage tiers or in the storage unit as a whole.

The capacity managing unit may be embodied as hardware or software or a combination of both and be implemented together with the monitoring unit in a dedicated processing unit. The capacity managing unit preferably is configured to select one or more of the data segments or copies thereof that may be considered as more suitable for erasure than others. Accordingly, the selection is taken dependent on at least a relevance metric indicating the value of each data segment, i.e. the relevance classes introduced before. It is preferred, that the data segments with the least relevance metric, i.e. the lowest relevance class indicating the lowest relevance of the corresponding data segment be suggested for erasure or at least for erasure of copies thereof. In one embodiment, the capacity managing unit takes a class-wise approach and, for example, suggests the data segments belonging to a common relevance class for erasure without differentiating between the data segments within such relevance class. In a different embodiment, however, the capacity managing unit takes an individual approach to data segments and may even differentiate between importance values of data segments with a common relevance class, e.g. by means of further evaluation of the content of the data segments, or by means of applying additional information available for the data segments.

The selection may be performed dependent on additional parameters, such as one or more of an access frequency to the subject data segments, an age of the data segments, a persistence metric assigned to the data segments, an obsolescence of data segments, etc.

In a preferred embodiment, the following metric is introduced for the storage capacity manager to determine which data segments are selected for further action, e.g. for deletion, suggestion for deletion, or a reduction or suggestion for reduction of redundancy.

$$R(d_\ell) = f_R(i_r(c_k), i'_p(c'_j), i_a(d_\ell), i_s(d_\ell))$$

or, in a more specific embodiment

$$R(d_\ell) = f_R(\gamma\, i_r(c_k) + \delta\, i'_p(c'_j) + \eta\, i_a(d_\ell) + \kappa\, i_s(d_\ell)), \qquad (7)$$

where

- $i_r(c_k)$ is a relevance class to which data segment d_ℓ belongs, i.e., $d_\ell \in c_k$, where the index ℓ denotes the data segment number;
- $i'_p(c'_j)$ is a popularity class c'_j to which data segment d_ℓ belongs, i.e., $d_\ell \in c'_j$;
- $i_a(d_\ell)$ is an age of the data segment d_ℓ ;
- $i_s(d_\ell)$ is a persistence of a data segment d_ℓ .

In general, the popularity class is different from the relevance class, and the popularity class of a data segment may vary with time. The popularity class may be defined as in connection with the storage system described above, and a determination of which may be supported by an access pattern evaluator such as described above. The relevance class may be determined by means of a classification unit such as described above, and may be stored as metadata together with the data segment in the storage unit. The age of the data segment may denote the age for which the data segment resides in the storage unit. The persistence of a data segment may in one embodiment be defined by a user or an administrator of the storage unit and specifically may take a value in the interval [0, 1], where persistence level 1 means "never delete", and 0 means "obsolete data".

In a preferred embodiment, the storage capacity manager applies the following rule:

- If $R(d_\ell) < thr1$ then delete data segment d_ℓ ;
- If $thr1 < R(d_\ell) < thr2$ then reduce a redundancy of data segment d_ℓ; note that the term $\kappa\ i_s(d_\ell)$ alone must be able to be $> thr1$ to avoid unintended deletion;
- If $thr2 < R(d_\ell)$ then keep data segment d_ℓ unmodified.

It is preferred that one or more copies of a data segments are suggested for removal first before removing the data segment as such, i.e. all copies thereof. Instead of or in addition to a suggestion or a removal of data segments or copies thereof, or in case the storage capacity manager determines that all existing data segments in the storage unit are still important, a recommendation may be made by the storage capacity manager to a user or an administrator as to expand the storage capacity of the storage unit or at least a tier of the storage unit in case of a multiple tier storage unit.

An automatic recommendation for a storage capacity expansion may be based on or more of:

- a computation of a capacity required to extend the present storage capacity by x%; or to serve storage requirements for the next y months, based on a historic capacity growth rate;
- a determination of a storage device mix based on one or more of a storage unit needs, a current storage tier utilization, a historic capacity growth rate per storage tier, etc.

<u>Advantages and Applications</u>

A storage system as suggested in various embodiments addresses the content of the data segments to be stored and preferably classifies the data segments in real-time. Preferably, each data segment to be stored is associated with a relevance index reflecting the assigned relevance class and a popularity index reflecting the access frequency for data segments of the same relevance class in the data storage. Based on at least this two-fold information, the storage system allows a full automatic selection of an appropriate level of protection for each data segment, and a full automatic selection of a storage tier a certain data segment is to be initially stored, all without human intervention.

A heterogeneous storage infrastructure, including e.g. solid-state drives, hard-disk drives, and tape systems, can efficiently be used. Performance, reliability, security, and storage efficiency at low operating cost and power consumption are achieved by evaluating the importance of the stored information for the purpose of, e.g., of unequal data protection, intelligent tiering, and eventually erasure of obsolete data.

As explained in the previous sections, in embodiments of the storage system different levels of protection are granted to data segments to be stored, depending on the relevance of the information contained. In one embodiment, it is assumed that data segments received for storage are classified by a classifier into one out of K+1 relevance classes, depending on their information content. Preferably, data segments with poor information content due to, e.g., calibration procedures or presence of interference, are assigned to Class 0, and preferably are discarded or stored at the lowest possible cost. Data segments in the remaining K classes are input to K different block encoders for error correcting codes. Each encoder may be characterized by parameters n_i and k_i, where k_i is the number of data symbols being encoded, and n_i is the total number of code symbols in the encoded block. Specifically, a multi-tiered storage system with seven relevance classes is considered, where data segments are assigned to the various relevance classes according to a binomial distribution with parameter p. Again, the data segments assigned to Class 0 are assumed to be irrelevant. The data segments in Classes 1 to 6 are then encoded with a RS $(64,k_i)$ code from $GF(2^8)$, where k_i goes from 60 to 40, i.e., the code length n is held at a constant value equal to 64, whereas the number of data symbols is given by $k_i = 64-4i$, i.e., the number of data symbols decreases from $k_1 = 60$ to $k_6 = 40$. The redundancy thus increases from 4 symbols within a codeword for Class 1 to 24 symbols for Class 6. To assess a gain in storage efficiency that is obtained by the assumed storage system, consider an application where the data segments correspond to images with 100x100 pixels. Data segments might be assigned to Class 0 and discarded if collected, e.g., during calibration of experiments or in the presence of interference. For a random channel bit-error probability of 10^{-3}, the six classes define sequences of images where in the average one pixel is in error every 1, 10^2, 10^5, 10^8, 10^{11}, 10^{14} images after retrieval, respectively. The efficiency gain obtained by the considered system with unequal error protection and binomial class probability distribution over a system that adopts RS encoding by a (64,40) code from GF(2^8) for all data segments, is given in percent by

$$g = \left(\frac{64}{40} \frac{1}{\sum_{k=1}^{6} \binom{6}{k} p^k (1-p)^{6-k} \frac{64}{64-k}} - 1 \right) \times 100$$

A storage system as introduced, which may also be referred to as cognitive data storage system, preferably is applied for big data applications. In such storage system, information may be efficiently extracted, stored, retrieved, and managed. Preferably, in a first step, online detection and classification techniques are applied on incoming data segments. In this step, the occurrence of events that are associated with valuable information are preferably detected and classified. Preferably, in a second step, the result of the classification procedure together with information about the access patterns of similarly classified data is used to determine with which level of protection against errors, and within which tier of the storage system the incoming data segments are to be initially stored.

For instance, this cognitive approach could be useful for application in an existing system (such as LOFAR) or in the future square kilometer array (SKA) telescope system. In particular, it may be applied to optimize future data access performance. Various workload characteristics may be evaluated for data placement optimization, such as sequentiality and frequency of subsequent data accesses. Based on this information, the appropriate tier for storing the data can be determined. Moreover, predictions regarding subsequent data accesses can enable effective caching and pre-fetching strategies.

In the specific embodiment of the square kilometer array (SKA), the functions of the classification unit may be performed by an enhanced version of a Science Data Processor (SDP). The SDP preferably has the task to automatically calibrate and image the incoming data, from which science oriented generic catalogues can be automatically formed prior to the archiving of images that are represented by the incoming data segments. Note that an event detector/classifier pair in the classification unit may face the challenging task of determining in real time a set of features related to a detected event, for example real time detection and machine-learned classification of variable stars from time-series data. In this case, the detection of variable stars using the least squares fitting of sinusoids with a floating mean and over a range of test frequencies, followed by tree-based classification of the detected stars may in one embodiment be well suited for online implementation. Within the current SKA architecture, the functions of a Multi-Tier Storage (MTS) system preferably are performed by an enhanced version of a Science Data Archive Facility.

For applications within the healthcare industry, the functions of the classification unit preferably depend on the context of the data being stored. For example, if data segments being collected are used for a cohort study, the parts that are relevant to the study may be classified as more important than other data. In the context of personalized medicine, medical records may be identified by their type, e.g., biochemistry, hematology, genomics, hospital records. Within each type, relevant features may be classified and associated with a certain level of importance.

Figure 1 shows a block diagram of a storage system according to an embodiment of the present invention. The storage system comprises two main subsystems, i.e. a classification unit 1 also referred to as Real Time Processing & Classification (RTPC), and a storage unit 2 also referred to as Multi-Tier Storage (MTS). In the classification unit 1, an incoming data stream containing data segments is elaborated by a real-time processing unit 15, typically to perform one or more of filtering operations, suppression of spurious data segments, e.g., removing interference in the context of astronomical data application, ensuring privacy of medical records by pseudonymization in the context of cohort studies in the healthcare industry, or extracting relevant information from medical records in the context of personalized medicine. An output of the real-time processing unit 15 is presented to a set of N online event detectors 11. Each of the N event detectors 11 determines whether the occurrence of an event, which may be associated with predefined information, is detected within a segment of the incoming data stream. Each event detector 11 of the set may be configured to detect a specific event that is different from the events the other event detectors 11 of the set are expected to detect.

In general, real-time classification may refer to any initial data evaluation that may take place while guaranteeing a predetermined sustained rate of the incoming data stream. Whenever a relevant event is detected by one of the N event detectors, an associate online classifier 12 assigns the data segment, which contains the information related to the event, to one of K+1 relevance classes with $K \geq 0$ depending, e.g., on the presence or absence of features that characterize the event. Data segments, where event-related information is not detected, are assigned by default to a Class 0. Note that a set of N buffers 13 is included in the data paths to compensate for delays introduced by the associate event detector 11 and classifier 12. Also note that in a preferred embodiment, several pairs of event detectors 11 and classifiers 12 may be operating in parallel, if events of different nature are deemed relevant, as illustrated in Fig. 1 for N detector 11/classifier 12 pairs, with N=3. In this case, a data segment may be associated with multiple tags assigned by the various classifiers 12, and a buffer 16 in the main data path to the storage unit 2 preferably is dimensioned to accommodate a largest delay expected to be introduced by the classifiers.

In other embodiments, only the occurrence of a single event is desired to be detected, in which case the event detector at the same time acts as classifier – or the classifier acts as event detector. In a different embodiment, the occurrence of an event may be known a priori and only event features are desired to be identified, in which case the event detector/s 11 is/are not needed.

The data segments and the related class information – which class information may be subsumed under the data segments metadata - are received by a selector 14, which has the task of determining with which level of protection and to which tier each data segment received is to be stored. This decision depends on the information on the relevance class and on an access pattern to this relevance class, which is obtained from an access pattern evaluator 24 assigned to the storage unit 2.

The storage unit 2 receives from the classification unit 1 the processed sequence of incoming data segments to be stored in a multi-tier storage 21 containing M storage tiers, together with the individual information for each data segment about detected events, and identified relevance classes and possibly other features. This information preferably is utilized to assign a protection level for the respective data segment and an initial placement in one of the available storage tiers 21. An M-tier storage system with J data segment protection levels, with M=3 and J=3, is illustrated for example in Fig. 1. The three storage tiers 21 might correspond to different type of storage media, e.g., SSDs, HDDs, and tape. For performance optimization during normal system operation, frequently accessed data or randomly accessed data may preferably be placed on SSDs, whereas less frequently accessed data or sequentially accessed data may be stored on HDDs or on tape.

Prior to being stored on the physical media corresponding to the selected storage tier 21, each data segment is presented to an encoder 23, which provides different levels of protection, for example using unequal error protection (UEP), depending on the relevance class information. In an embodiment, compression and/or deduplication of the data segments may be considered in addition to UEP. A data segment with a high relevance class preferably is associated with a high value. Its information content is such that a loss would be associated with a high cost, and therefore the data segment is protected with a higher level of redundancy. The required level of redundancy may be provided by error correction coding or erasure coding, by storing replicas of the data segments, or by a combination of these techniques. Note that compression and/or deduplication of the data segments may be considered in addition to UEP.

An access pattern evaluator 24 of the storage unit 2 provides additional information about the popularity of the data segments associated with a certain relevance class. Every time a data segment is accessed in the storage unit 2, the associated metadata information including the class information is provided to the access pattern evaluator, which learns about the popularity of the information content in the data segments from the way they are being accessed. Access patterns may be found at various levels, e.g., activity during various times of a day, sequence of reads and writes, access sequentiality, and number of users retrieving the data. This information is used to further classify data segments into one of several popularity classes.

Subsequently, the access pattern evaluator 24 sends information to the selector 14 in the classification unit 1, which accordingly updates a metric for initial decision on level of protection and storage tier of individual data segments. Therefore, the selector 14 updates the criterion for initial data placement based on both data relevance classification and data access statistics. In this manner, a data segment that belongs to a certain relevance class is passed out to the storage medium and is protected against errors with a redundancy level that are most appropriate at a particular point in time. Following an initial data placement, the access pattern evaluator 24 monitors all data segments in the storage tiers 21 and places each in the appropriate popularity class.

In the present embodiment of Figure 1, a migrator 25 is provided, also referred to as storage relocation manager, which is preferably arranged in the storage unit 2. The migrator 25 receives information from the access pattern evaluator 24, and as such receives access frequencies to the individual classes. This information enables the migrator 25 to place data segments in the right storage tier 21 to enhance access performance. Specifically, the migrator 25 has the task of moving data segments stored in one storage tier 21 to another storage tier 21 if such movement is indicated by the present access patterns to such data segments, and specifically by the access patterns of the class to which the respective data segment belongs. This classification enables the migrator 25 to place the stored data segments in the right storage tier to enhance access performance.

In the present embodiment of Figure 1, storage efficiency and access performance are further optimized by continuously monitoring the access patterns and updating the criteria for data segment placement in a storage capacity manager 26. When the amount of data segments stored in the storage unit system approaches a system capacity, an automatic selection of data segments for deletion or deletion of copies thereof is provided by the storage capacity manager 26. The selected data segments and / or copies thereof may be deleted and / or may be suggested for deletion and / or a system capacity expansion is requested, all preferably based on the importance and the access patterns of the data segments stored in the system.

FIG. 2 illustrates a flowchart of a method for storing a data segment in a storage tier of a storage unit comprising at least two storage tiers, according to an embodiment of the present invention. In step 21 a data segment is received for storage. In step 22, at least one out of a set of at least two relevance classes is assigned to the data segment dependent on information comprised in the data segment. In step 23, a level of protection is determined for the classified data segment dependent on the relevance class assigned. In step 24, a storage tier is determined out of the at least two storage tiers for storing the classified data segment to dependent on at least access frequency information received for data segments in the same relevance class, and dependent on the characteristics of the storage tiers available. In step 25, a redundancy level is determined for the data segment dependent on the determined protection level and the determined storage tier. In step 26, the classified data segment including the assigned relevance class is stored in the determined storage tier and copies of the data segment are stored in the determined storage tier or in a different storage tier according to the determined redundancy level.

FIG. 3 illustrates a flowchart of a method for relocating a data segment presently stored in a storage tier of a storage unit with at least two storage tiers, according to an embodiment of the present invention. In step 31, an access frequency to a relevance class is monitored. In step 32, it is verified if a change in an access frequency Δaf to data segments stored in the storage unit and belonging to the subject relevance class exceeds a threshold t1. If not, the access frequency is continued to be monitored in step 31. If yes (y), a data segment assigned to the subject relevance class and stored in a present storage tier is moved to a different storage tier, i.e. a target storage tier. In step 34, protection measures are adapted for this relocated data segment in order to achieve a protection level assigned to the data segment.

FIG. 4 illustrates a method for managing a capacity of a storage unit for storing data segments, according to an embodiment of the present invention, preferably executed by a storage capacity manager according to an embodiment of the present invention. In step 41 a remaining capacity C of the storage unit is monitored. In step 42 it is verified, if the remaining storage capacity C is less than a threshold thr1. If not, the remaining capacity C is continued to be monitored. If yes (y), data segments are deleted in step 43 thereby reducing the occupied storage space in the storage unit. Step 43 includes the selection of storage segments to be deleted. In the present example, the data segments are selected dependent on a relevance class assigned, and the data segments assigned to the lowest relevance class are removed. In step 44, it is verified if the remaining storage capacity C is less than a threshold thr2, which preferably is less than the threshold thr1. If not, the remaining capacity C is continued to be monitored. If yes (y), a storage capacity expansion is recommended to an administrator of the storage unit.

The present invention may be a system, a method, and/or a computer program product. The computer program product may include a computer readable storage medium (or media) having computer readable program instructions thereon for causing a processor to carry out aspects of the present invention.

The computer readable storage medium can be a tangible device that can retain and store instructions for use by an instruction execution device. The computer readable storage medium may be, for example, but is not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. A non-exhaustive list of more specific examples of the computer readable storage medium includes the following: a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), a static random access memory (SRAM), a portable compact disc read-only memory (CD-ROM), a digital versatile disk (DVD), a memory stick, a floppy disk, a mechanically encoded device such as punch-cards or raised structures in a groove having instructions recorded thereon, and any suitable combination of the foregoing. A computer readable storage medium, as used herein, is not to be construed as being transitory signals *per se*, such as radio waves or other freely propagating electromagnetic waves, electromagnetic waves propagating through a waveguide or other transmission media (e.g., light pulses passing through a fiber-optic cable), or electrical signals transmitted through a wire.

Computer readable program instructions described herein can be downloaded to respective computing/processing devices from a computer readable storage medium or to an external computer or external storage device via a network, for example, the Internet, a local area network, a wide area network and/or a wireless network. The network may comprise copper transmission cables, optical transmission fibers, wireless transmission, routers, firewalls, switches, gateway computers and/or edge servers. A network adapter card or network interface in each computing/processing device receives computer readable program instructions from the network and forwards the computer readable program instructions for storage in a computer readable storage medium within the respective computing/processing device.

Computer readable program instructions for carrying out operations of the present invention may be assembler instructions, instruction-set-architecture (ISA) instructions, machine instructions, machine dependent instructions, microcode, firmware instructions, state-setting data, or either source code or object code written in any combination of one or more programming languages, including an object oriented programming language such as Smalltalk, C++ or the like, and conventional procedural programming languages, such as the

"C" programming language or similar programming languages. The computer readable program instructions may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider). In some embodiments, electronic circuitry including, for example, programmable logic circuitry, field-programmable gate arrays (FPGA), or programmable logic arrays (PLA) may execute the computer readable program instructions by utilizing state information of the computer readable program instructions to personalize the electronic circuitry, in order to perform aspects of the present invention.

Aspects of the present invention are described herein with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems), and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer readable program instructions.

These computer readable program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks. These computer readable program instructions may also be stored in a computer readable storage medium that can direct a computer, a programmable data processing apparatus, and/or other devices to function in a particular manner, such that the computer readable storage medium having instructions stored therein comprises an article of manufacture including instructions which implement aspects of the function/act specified in the flowchart and/or block diagram block or blocks.

The computer readable program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other device to cause a series of operational steps to be performed on the computer, other programmable apparatus or other device to produce a computer implemented process, such that the instructions which execute on the computer, other programmable apparatus, or other device implement the functions/acts specified in the flowchart and/or block diagram block or blocks.

The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods, and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of instructions, which comprises one or more executable instructions for implementing the specified logical function(s). In some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts or carry out combinations of special purpose hardware and computer instructions.

CLAIMS

1. A storage system, comprising:

- a storage unit (2) comprising at least two storage tiers (21M),
- an access pattern evaluator (24) configured to provide information about a frequency at which data segments stored in one or more of the at least two storage tiers (21M) are accessed,
- a classification unit (1) configured to assign at least one out of a set of at least two relevance classes to a data segment received for storing in the storage unit (2) dependent on information comprised in the data segment,
- a selector (14) configured to determine a storage tier (21m) out of the at least two storage tiers (21M) for storing the classified data segment to, dependent on at least access frequency information provided by the access pattern evaluator (24) for data segments in the same relevance class, and configured to determine a level of protection for the classified data segment dependent on at least the relevance class assigned, and
- logic (22) for storing the classified data segment including the assigned relevance class to the determined storage tier (21i) and according to the determined level of protection.

2. The storage system according to claim 1,

- wherein the classification unit (1) comprises a set of event detectors (11N) with each event detector (11n) of the set being configured to detect a different event in the data segment received for storing, and
- wherein the classification unit (1) comprises a set of classifiers (12N), each classifier being assigned to a different one of the event detectors (11n), and each classifier (12n) of the set being configured to assign at least one out of one or more relevance classes pre-selected from the set of relevance classes for the event to be detected by the assigned event detector (11n).

3. The storage system according to claim 1 or claim 2, comprising
a real-time data processing unit (15) for real-time processing of input data segments and providing a sequence of data segments for storing.

4. The storage system according to any one of the preceding claims,

- wherein the selector (14) is configured to determine the storage tier (21m) for the classified data segment dependent on the access frequency information provided by the access pattern evaluator (24) for data segments in the same relevance class, and dependent on the relevance class assigned, and
- wherein the selector (14) is configured to determine the level of protection for the classified data segment dependent on the relevance class assigned, and dependent on the access frequency information provided by the access pattern evaluator (24) for data segments in the same relevance class.

5. The storage system according to any one of the preceding claims,

- wherein the selector (14) is configured to determine a redundancy level for the classified data segment dependent on the determined protection level and the determined storage tier, the redundancy level specifying a number of copies of the classified data segment to be stored in the storage unit, and in particular specifying a number of copies of the classified data segment to be stored in which storage tier of the storage unit, and
- wherein the logic is configured to store copies of the classified data segment in the one or more storage tiers according to the determined level of redundancy.

6. The storage system according to any one of the preceding claims,

- wherein the selector (14) is configured to determine one or more of an error correction code and an erasure code to be applied to the classified data segment dependent on the determined protection level and dependent on the determined storage tier, and
- wherein the logic is configured to store the classified data segment with the one or more of the determined error correction code and determined erasure code in the storage unit.

7. Method for storing a data segment in a storage tier of a storage unit comprising at least two storage tiers, comprising:

- assigning at least one out of a set of at least two relevance classes to the data segment dependent on information comprised in the data segment,
- receiving information about a frequency at which data segments stored in one or more of the at least two storage tiers are accessed,

- determining a storage tier (21m) out of the at least two storage tiers (21M) for storing the classified data segment to dependent on at least access frequency information received for data segments in the same relevance class,
- determining a level of protection for the classified data segment dependent on at least the relevance class assigned, and
- storing the classified data segment including the assigned relevance class to the determined storage tier (21i) and according to the determined level of protection.

8. A storage tier relocation manager for relocating a data segment presently stored in a storage tier of a storage unit (2) with at least two storage tiers, and in particular stored in a storage system according to any one of the preceding claims 1 to 6, the data segment having assigned a protection level out of a set of protection levels, the storage tier relocation manager being configured to

- determine a target storage tier for the data segment dependent on access frequency information received for one or more of the data segment or a relevance class the data segment is assigned to,
- relocate the data segment to the target storage tier if the target storage tier is different from the present storage tier, and
- in case of a relocation apply a protection measure suitable for achieving the assigned protection level.

9. The storage tier relocation manager according to claim 8, being configured to

- in case of a relocation apply the protection measure by selecting one or more of:
    - a redundancy level specifying a number of copies of the data segment to be stored in one or more of the target storage tier and a non-target storage tier;
    - an error correction code to be applied to the data segment;
    - and an erasure code to be applied to the data segment; and
- store the data segment in the storage unit according to the selected one or more of the redundancy level, the error correction code and the erasure code.

10. Method for relocating a data segment presently stored in a storage tier of a storage unit (2) with at least two storage tiers, and in particular stored in a storage system according to any one of the preceding claims 1 to 6, the data segment having assigned a protection level out of a set of protection levels, comprising

- determining a target storage tier for the data segment dependent on access frequency information received for one or more of the data segment or a relevance class the data segment is assigned to,
- relocating the data segment to the target storage tier if the target storage tier is different from the present storage tier, and
- in case of a relocation applying a protection measure suitable for achieving the assigned protection level.

11. A storage capacity manager for a storage system comprising a storage unit (2) for storing data segments having at least one relevance class assigned per data segment, and in particular for a storage system according to any one of the preceding claims 1 to 6, the storage capacity manager comprising

– a monitoring unit for determining if a utilization of the storage unit fulfils a criterion,
– a capacity managing unit for, in response to the utilization of the storage unit fulfilling the criterion, selecting at least one data segment stored in the storage unit for one of a deletion thereof or a deletion of a copy thereof at least dependent on the at least one relevance class assigned.

12. The storage capacity manager according to claim 11,

- wherein the storage capacity manager is configured to manage a storage unit (2) comprising at least two storage tiers (21M),
- wherein the monitoring unit is configured to determine if the utilization of at least one of the at least two storage tiers (21M) fulfils the criterion, and
- wherein the capacity managing unit is configured to, in response to the utilization of at least one storage tier fulfilling the criterion, selecting the at least one data segment stored in this storage tier for one of a deletion thereof or a deletion of a copy thereof in the same storage tier or in a different storage tier.

13. The storage capacity manager according to claim 11 or claim 12,

- wherein the monitoring unit is configured to determine if the utilization of the storage unit falls below a capacity threshold, and
- wherein the capacity managing unit is configured to select the at least one data segment in response to the utilization of the storage unit falling below the capacity threshold.

14. The storage capacity manager according to any one of the preceding claims 11 to 13, wherein the capacity managing unit is configured to one of:

- suggest the at least one selected data segment for deletion or deletion of a copy thereof;
- suggest the at least one selected data segment for deletion or deletion of a copy thereof and delete the at least one selected data segment or a copy thereof in response to a user confirmation;
- delete the at least one selected segment or copies thereof.

15. The storage capacity manager according to any one of the preceding claims 11 to 14,

- wherein the capacity managing unit is configured to assign a retention class out of a set of retention classes to the data segments in the storage unit, each retention class out of the set indicating a measure for retaining the assigned data segment in the storage unit, the assignment of the retention class to a data segment being made dependent on the relevance class assigned to the data segment, and
- wherein the capacity managing unit is configured to select the at least one data segment for one of a deletion thereof or a deletion of a copy thereof dependent on the retention class assigned.

16. The storage capacity manager according to claim 15, wherein the capacity managing unit is configured to select the at least one data segment for deletion if the corresponding data segment is within a number of n data segments showing the lowest retention classes assigned.

17. The storage capacity manager according to claim 15 or claim 16,

- wherein the capacity managing unit is configured to manage the storage of data segments in a variable number of copies, and
- wherein the capacity managing unit is configured to select the at least one data segment for deleting one or more copies thereof if the corresponding retention class is below a retention threshold.

18. The storage capacity manager according to any one of the preceding claims 15 to 17, wherein the capacity managing unit is configured to determine the retention class to assign to a data segment in addition dependent on one or more of:

- an age of the data segment;
- access frequency information for the data segment or for the relevance class the data segment is assigned to;
- a persistence index;
- a storage capacity available in one or more of the other storage tiers in case of a tiered storage unit.

19. Method for managing a capacity of a storage unit for storing data segments having at least one relevance class assigned per data segment, and in particular for a storage system according to any one of the preceding claims 1 to 6, comprising

- determining if a utilization of the storage unit fulfils a criterion,
- in response to the utilization of the storage unit fulfilling the criterion, selecting at least one data segment stored in the storage unit for one of a deletion thereof or a deletion of a copy thereof at least dependent on the relevance class assigned.

20. Computer program product comprising a computer readable medium having computer readable program code embodied therewith, the computer readable program code comprising computer readable program code configured to perform a method according to any one of the preceding claims 7, 10, 19.

Intellectual Property Office

Application No:	GB1415248.2	**Examiner:**	Matthew Veale
Claims searched:	1-20	**Date of search:**	26 February 2015

Patents Act 1977: Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
A	-	US8315995 B1 (PEER FUSION INC)
A	-	US8370597 B1 (AMERICAN MEGATRENDS INC)
A,E	-	US2014/281322 A1 (SILICON GRAPHICS INT CORP)
A	-	US2012/117328 A1 (LSI CORP)

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC^{X} :

Worldwide search of patent documents classified in the following areas of the IPC

G06F

The following online and other databases have been used in the preparation of this search report

EPODOC, WPI, TXTE

International Classification:

Subclass	Subgroup	Valid From
G06F	0003/06	01/01/2006