

(51) International Patent Classification:
G06F 3/06 (2006.01) **G06F 11/34** (2006.01)(21) International Application Number:
PCT/JP2009/000504(22) International Filing Date:
9 February 2009 (09.02.2009)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **HITACHI, LTD.** [JP/JP]; 6-6, Marunouchi 1-chome, Chiyoda-ku, Tokyo, 1008280 (JP).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TANAKA, Toru** [JP/JP]; c/o Hitachi, Ltd., Systems Development Laboratory 1099, Ozenji, Asao-ku, Kawasaki-shi, Kanagawa, 2150013 (JP). **TAGUCHI, Yuichi** [JP/JP]; c/o Hitachi, Ltd., Systems Development Laboratory 1099, Ozenji, Asao-ku, Kawasaki-shi, Kanagawa, 2150013 (JP). **NAKAJIMA, Noriko** [JP/JP]; c/o Hitachi, Ltd., Systems Development Laboratory 1099, Ozenji, Asao-ku, Kawasaki-shi, Kanagawa, 2150013 (JP).(74) Agent: **HIRAKI, Yusuke**; Kamiya-cho MT Bldg. 19F, 3-20, Toranomon 4-chome, Minato-ku, Tokyo, 1050001 (JP).

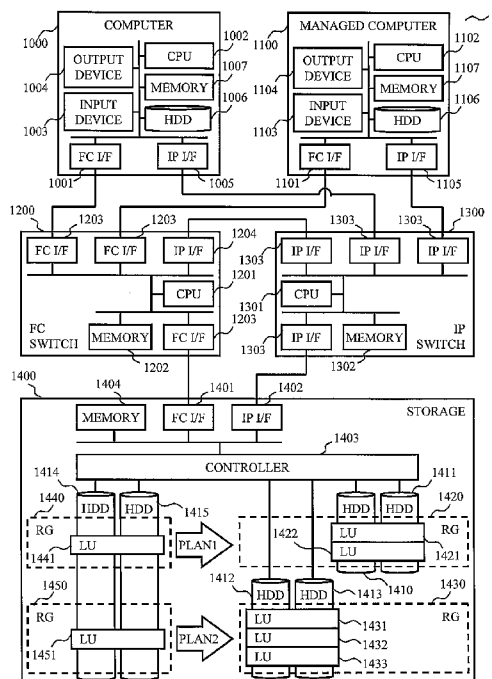
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: STORAGE SYSTEM

FIG. 1



(57) Abstract: The threshold excess is judged by a peak value, and therefore, since the threshold excess is judged by a value of high-workload IO of backup or the like that is designed on the assumption that the workload of storage is high, whether the performance limit has been reached remains uncertain. Furthermore, although an optimistic prediction trend with low workload trend and a pessimistic prediction trend with high pessimistic workload trend are presented, LUs that require improvement remain uncertain. Moreover, the time to improve the configuration and the effect of the improvement remain uncertain, and therefore, an improved configuration cannot be determined. Configuration information, performance information, and operation information are acquired from a storage, a performance limit excess time is predicted based on the performance information excluding LUs exceeding the performance trend, and an optimal improved configuration is determined from the performance limit after improvement and the time required for the improvement.



Published:

— *with international search report (Art. 21(3))*

Description

STORAGE SYSTEM

Technical Field

[0001] The present invention relates to a storage system, and for example, to a technique for managing the performance of a storage system.

Background Art

[0002] In recent years, the amount of data handled by a computer has increased as a result of the improvement in the performance of computer and the improvement in the speed of Internet connection. Since there is a limit to the performance of storage that processes data handled by a computer, proactive performance management for improving the configuration of storage before exceeding the performance limit is highly expected. Therefore, a technique for monitoring the workload of storage and issuing a warning when the threshold is exceeded is disclosed (see Patent Document 1). Additionally, software for calculating the future prediction from the workload trend of storage is disclosed (see Non-Patent Document 1).

Patent Citation 1: JP Patent Publication (Kokai) No. 2004-072135.

Non Patent Citation 1: JP1/HiCommand Tuning Manager, Reference Guide, pp.114-130

Disclosure of Invention

Technical Problem

[0003] However, according to the technique of Patent Document 1, the limit is judged based on whether the instantaneous peak value of workload exceeds the threshold. Therefore, since the threshold excess is judged by the value of high-workload IO of backup or the like that is designed on the assumption that the workload of storage is high, whether the performance limit has been reached remains uncertain.

[0004] Furthermore, although the technique of Non-Patent Document 1 presents an optimistic prediction trend with low workload trend and a pessimistic prediction trend with high pessimistic workload trend, LUs that require improvement remain uncertain. Moreover, according to a combination of Patent Document 1 and Non-Patent Document 1, the time to improve the configuration (although determined by time for changing the configuration and the logical processing time in the computer, what is dominant is the processing time by the computer for moving the data in a configuration with improved (extended) required data (thus, the latter)) and the effect of the improvement remains uncertain. Therefore, an improved configuration cannot be determined.

[0005] The present invention has been made in view of the forgoing circumstances, and the

present invention provides a technique that accurately predicts a performance trend of storage and a performance limit excess time without being affected by an instantaneous peak value of workload and that proposes an optimal improved configuration.

Technical Solution

- [0006] To solve the problems, the present invention acquires configuration information, performance information, and operation information from a storage, predicts the performance limit excess time based on the performance information excluding LUs exceeding a performance trend, and determines an optimal improved configuration from the performance limit after improvement and the time required for the improvement.
- [0007] More specifically, the present invention relates to a storage system (1) comprising a storage (1400) and a performance managing unit (1100) that manages processing performance of the storage. The performance managing unit obtains an approximate function (approximate straight line) of an amount of IO based on the transition of the amount of IO from a computer (1000) to the storage within a predetermined period to set the approximate function as a performance trend (7001) of the storage, further calculates a standard deviation of the amount of IO, and outputs the performance trend and comparison information between the standard deviation and the performance limit of the storage to an output unit (16004). The performance managing unit may be designed to calculate the performance trend and the standard deviation every time a computation instruction (pressing of an OK button (see Figure 16)) is inputted by the administrator and to output information of comparison with the performance limit of the storage to the output unit.
- [0008] The performance managing unit sets a minimum value among a total value of performance limits of disk devices (1410 to 1413) in each RAID group in the storage, a performance limit of a controller (1403) of the storage in each RAID group, and a performance limit of a bandwidth (host path 1401) from the computer to the storage in each RAID group as a performance limit of the storage. More specifically, the performance managing unit calculates an IO pattern for each RAID group, the IO pattern including an average of read/write rate and an average of read hit rate, and refers to a performance limit table storing the performance limits of the disk devices, the controller, and the host path corresponding to combinations of configurations of the IO patterns and the RAID groups to acquire the performance limits of the disk devices, the controller, and the host path from configurations of the calculated IO patterns and RAID groups to be processed.
- [0009] The performance managing unit further reads out a prepared extension plan of the disk devices in the storage, calculates an extension time in the case where the con-

figuration of the extension plan is implemented, the extension time including the time required for moving data to the extended disk devices, and outputs the comparison information as well as the extension plan and the extension time (16003) to the output unit.

[0010] The performance managing unit may be designed to further calculate a time zone (21003), in which a job is executed, from execution schedule information (17001) of the job in the computer or to calculate the performance trend from an amount of IO of another time zone to calculate the approximate function. In this case, the performance managing unit takes the sum of a plurality of job execution periods to specify the time zone in which the job is executed.

[0011] The performance managing unit may also be designed to narrow down the targets to RAID groups including storage areas having an amount of IO exceeding the value of the performance trend obtained first plus the standard error and to recalculate the performance trend of the targets.

[0012] Further features of the present invention will become apparent from the best mode for carrying out the present invention and the appended drawings.

Advantageous Effects

[0013] According to the present invention, a performance trend and a performance limit excess time of storage can be accurately predicted without being affected by a peak value, and an optimal improved configuration can be proposed. Therefore, proactive performance management can be realized.

Brief Description of the Drawings

[0014] [fig.1]Figure 1 is a diagram of a schematic configuration of a storage system according to embodiments of the present invention.

[fig.2]Figure 2 is a diagram of an internal configuration of a memory of a computer.

[fig.3]Figure 3 is a diagram of an internal configuration of a memory of a managed computer.

[fig.4]Figure 4 is a diagram of an internal configuration of a memory of a storage.

[fig.5]Figure 5 is a diagram of the contents of a computer setting table in the computer.

[fig.6]Figure 6 is a diagram of the contents of a performance limit information table in the managed computer.

[fig.7]Figure 7 is a diagram of the contents of a performance trend information table in the managed computer.

[fig.8]Figure 8 is a diagram of the contents of an improved configuration information table in the managed computer.

[fig.9]Figure 9 is a diagram of the contents of a storage configuration information table in the storage.

[fig.10]Figure 10 is a diagram of the contents of a storage performance information table in the storage.

[fig.11]Figure 11 is a flow chart for explaining a performance management process in an embodiment of the present invention.

[fig.12]Figure 12 is a flow chart for explaining details of a performance analysis process constituting part of the performance management process.

[fig.13]Figure 13 is a flow chart for explaining details of a performance limit calculation process of each component in the performance analysis process (Figure 12).

[fig.14]Figure 14 is a flow chart for explaining details of a performance trend calculation process in the performance analysis process (Figure 12).

[fig.15]Figure 15 is a flow chart for explaining details of an improved configuration determination process constituting part of the performance management process.

[fig.16]Figure 16 is a diagram of a performance management screen example in an embodiment of the present invention.

[fig.17]Figure 17 is a diagram of an internal configuration of a memory of the computer in a second embodiment of the present invention.

[fig.18]Figure 18 is a diagram of the contents of an operation information table in the computer.

[fig.19]Figure 19 is a flow chart for explaining a performance management process in the second embodiment of the present invention.

[fig.20]Figure 20 is a flow chart for explaining details of a narrowing-down process in the performance management process (Figure 19).

[fig.21]Figure 21 is a diagram of a registration table for explaining the concept of a job execution period and a job execution time zone.

Explanation of Reference

[0015] 1000 ... computer, 1100 ... managed computer, 1200 ... FC switch, 1300 ... IP switch, 1400 ... storage, 16000 ... output screen

Best Mode for Carrying Out the Invention

[0016] The present invention predicts a realistic workload excess time (limit time) in a storage system from the past usage without being bothered by an instantaneous peak value of workload and provides an effective configuration improvement plan based on the prediction result.

[0017] The embodiments of the present invention will now be described with reference to the appended drawings. It should be noted that the present embodiments only illustrate an example for realizing the present invention and should not be construed as limitations of the technical scope of the present invention. Common configurations in the drawings are designated with the same reference numerals.

[0018] 1) First Embodiment

<Configuration of Storage System>

Figure 1 is a diagram of a configuration (common configuration in the embodiments) of a storage system 1 in the embodiments of the present invention. In the present invention, the storage system 1 is implemented to calculate the performance limit of a storage 1400 used by a computer 1000.

[0019] The storage system 1 includes at least one computer 1000, at least one managed computer 1100, at least one FC switch 1200, at least one IP switch 1300, and at least one storage 1400.

[0020] The computer 1000 is a computer for performing input and output to the storage 1400 and comprises: an FC_I/F 1001 that transmits and receives input/output data to and from the storage 1400; an IP_I/F 1005 that transmits and receives management data to and from the managed computer 1100; a CPU 1002 that executes programs to control the entire computer; a memory 1007 that is a storage area of the programs; a storage 1006 that stores the programs, user data, or the like; an input device 1003 such as a keyboard and a mouse for inputting information from the user; and an output device 1004 such as a display that displays information to the user.

[0021] The managed computer 1100 is a computer that manages the computer 1000 and the storage 1400 and comprises: an FC_I/F 1101 that transmits and receives input/output data and control data to and from the storage 1400; an IP_I/F 1105 that transmits and receives management data to and from the computer 1000 and the storage 1400; a CPU 1102 that executes programs to control the entire computer; a memory 1107 that is a storage area of the programs; a storage 1106 that stores the programs, user data, and the like; an input device 1103 such as a keyboard or a mouse for inputting information from the user; and an output device 1104 such as a display that displays information to the user.

[0022] The FC switch 1200 is a switch device for transferring the input/output data from the managed computer 1000 to the storage 1400 or the like and comprises: an FC_I/F 1203 that transmits and receives the input/output data; an IP_I/F 1204 that transmits and receives management data; a CPU 1201 that executes programs to control the entire FC switch; and a memory 1202 that is a storage area of the programs and data.

[0023] The IP switch 1300 is a switch device that transfers management data from the managed computer 1100 to the computer 1000 or the like and comprises: an IP_I/F 1303 that transmits and receives the management data; a CPU 1301 that executes programs to control the entire IP switch; and a memory 1302 that is a storage area of the programs and data.

[0024] The storage 1400 is a node that processes the input/output data from the computer 1000 and comprises: an FC_IF 1401 that receives the input/output data transferred

from an FC switch; an IP_I/F 1402 that receives the management data from the managed computer 1100; a controller 1403 that executes programs to control the entire storage; a memory 1404 that is a storage area of the programs; and a disk device that saves user data. Disk devices 1410, 1411, 1412, and 1413, RAID groups 1420 and 1430 in which the disk devices are logically partitioned, LUs 1421, 1422, 1431, 1432, and 1433 in which the RAID groups are partitioned in a form visible by the user denote a storage configuration before extending the storage. Storages 1414 and 1415, RAID groups 1440 and 1450, and LUs 1441 and 1451 denote PLANs 1 and 2 for extension (extension plans). Thus, the PLANs 1 and 2 are not yet connected to the controller 1403.

[0025] <Internal Configuration of Memory>

Figure 2 is a diagram of a memory configuration of the computer 1000. Upon startup, the computer 1000 reads out, from the HDD 1006, for example, a data input/output program 2001 for inputting and outputting data to the storage 1400, a computer setting management program 2002 for managing setting information of the computer, and a computer setting table 2003 that is setting information of the computer and stores them in the memory 1007.

[0026] Figure 3 is a diagram of a memory configuration of the managed computer 1100. Upon startup, the managed computer 1100 reads out, from the HDD 1106 for example, a performance management program 3001 for managing the performance of the storage 1400, a performance limit calculation program 3002 that is a program for calculating the performance limit of the storage 1400, a performance trend calculation program 3003 that is a program for calculating the performance trend of the storage 1400; an improved configuration determination program 3004 that is a program for determining an improved configuration of the storage 1400, a performance limit information table 3005 that is performance limit information of the storage 1400, a performance trend information table 3006 that is performance trend information of the storage 1400, an improved configuration information table 3007 that is improved configuration information of the storage 1400 and stores them in the memory 1107.

[0027] Figure 4 is a diagram of a memory configuration of the storage 1400. Upon startup, the storage 1400 reads out, from any HDD for example, a data processing program 4001 for causing the computer 1000 or the like to access the storage 1400, a storage information management program 4002 for managing configuration information and performance information of storage, a configuration information table 4003 that is configuration information of storage, and a storage performance information table 4004 that is performance information of storage and stores them in the memory 1404.

[0028] Although the tables and the programs are read out from the target disk devices (HDD) herein, the example is not limited to this. The tables and the programs may be

read out from separate storage means (for example, service device).

[0029] <Contents of Read Out Tables>

Figure 5 is a diagram of a configuration of the computer setting table 2003. The computer setting table 2003 comprises COMPUTER 5001 that is an identifier of computers, LU 5002 that is an identifier of LUs used by the computers, TARGET_IF 5003 denoting connection points of the LUs used by the computers, and JOB 5004 denoting job types of the LUs used by the computers. Although the applications of the LUs are predetermined as shown in JOB 5004, the applications may be able to be changed later.

[0030] Figure 6 is a diagram of a configuration of the performance limit information table 3005. The performance limit information table 3005 comprises STORAGE 6001 that is an identifier of storages, RG 6002 that is an identifier of RAID groups, CONTROLLER 6003 denoting performance limit values of controllers, HDD 6004 denoting performance limit values of HDDs, and HOST_PATH 6005 denoting performance limit values of paths from the computers to the storages.

[0031] In Figure 6, the values inputted to the CONTOROLLER 6003, the HDD 6004, and the HOST_PATH 6005 denote the numbers of IO that can be processed (IOPS). The values calculated by the performance limit calculation program 3002 in advance are inputted to the fields.

[0032] Figure 7 is a diagram of a configuration of the performance trend information table 3006. The performance trend information table 3006 comprises TREND 7001 denoting a formula of the performance trend of storage, OPTIMISTIC_TREND 7002 denoting a formula of an optimistic performance trend in which the workload to the storage is predicted low, and PESSIMISTIC_TREND 7003 denoting a formula of a pessimistic performance trend in which the workload to storage is predicted high. These are obtained by using data of workload from now (NOW) to a past predetermined period to perform computations (Formulas 1 to 6) described below, every time the user presses an OK button (see Figure 16). As the workload constantly changes depending on the usage of the storage system 1, the formulas denoting the performance trend inputted to Figure 7 change every time the button is pressed. Although the formulas indicating the performance trend are obtained every time the OK button is pressed, the formulas may be sequentially obtained according to the passed time.

[0033] Figure 8 is a diagram of a configuration of the improved configuration information table 3007 (thus, improvement plans). The improved configuration information table 3007 comprises STORAGE 8001 that is an identifier of storages, HDD 8002 that is an identifier of HDDs to be extended, RG 8003 that is an identifier of RAID groups to be extended, RAID_LEVEL 8004 denoting RAID configuration types of the RAID groups to be extended, LU 8005 that is an identifier of LUs to be extended, SIZE 8006

denoting the capacities of the LUs to be extended, LIMIT 8007 denoting performance limits after extension, and TIME 8008 denoting time required for the extension. Thus, the table illustrates, for example, a PLAN for adding HDDs, which are shown in 8002 and to which the RAID 8003, the RAID_LEVEL 8004, and the LU 8005 are set, to the storages shown in 8001. The LIMIT 8007 and the TIME 8008 are blank at the initial stage.

[0034] Figure 9 is a diagram of a configuration of the storage configuration information table 4003. The storage configuration information table 4003 comprises STORAGE 9001 that is an identifier of storages, RG 9002 that is an identifier of RAID groups, RAID_LEVEL 9003 denoting RAID configuration types of the RAID groups, LU 9004 that is an identifier of LUs, and SIZE 9005 denoting capacities of the LUs. In the present description, information of each row will be called one record.

[0035] Figure 10 is a diagram of a configuration of the storage performance information table 4004. The storage performance information table 4004 comprises TIME 10001 denoting acquisition date and time of performance information, STORAGE 10002 that is an identifier of storages, RG 10003 that is an identifier of RAID groups, LU 10004 that is an identifier of LUs, READ_IOPS 10005 denoting read amounts of IO per second received by the storages, WRITE_IOPS 10006 denoting written amounts of IO per second received by the storages, and HIT 10007 denoting cache hit ratios of the read IO. The cache hit ratio herein denotes the rate of the number of times the processes have completed (returned data to the host computer) after being hit by cache without accessing up to the HDD in the READ_IO (apparent workload) (apparent workload x cache hit ratio = substantial workload of HDD). Thus, the completion of process by cache means that the workload is not imposed on the HDD, and the performance limit of the HDD differs depending on the value. Therefore, setting the cache hit ratio as a parameter for acquiring the performance limit is effective.

[0036] In the present embodiment, there is no record in the performance limit information table 3005, the performance trend information table 3006, and the improved configuration information table 3007.

[0037] <Performance Management Process>

Figure 11 is a flow chart for explaining a performance management process of the present embodiment. The administrator uses, for example, the computer 1000 to transmit a display request of a performance management screen (see Figure 16: PLAN, performance graph, and the like are not displayed at the initial stage) to the managed computer 1100 (step S11001). The performance management program (performance management processor) 3001 of the managed computer 1100 that has received the request transmits a performance management screen 2100 to the computer 1000 used by the administrator (step S11002). Although the computer used by the administrator is

the computer 1000 herein, the computer may be the managed computer 1100 or may be other computers.

[0038] The computer 1000 used by the administrator then transmits a request of performance limit calculation and improved configuration display to the managed computer 1100 (step S11003). The performance management program (performance management processor) 3001 of the managed computer 1100 that has received the request transmits a transmission request of the storage configuration information table 4003 and the storage performance information table 4004 to the storage 1400 (step S11004). The storage information management program 4002 of the storage 1400 that has received the request transmits the storage configuration information table 4003 and the storage performance information table 4004 to the managed computer 1100 (step S11005). The timing of transmission may be the timing of reception of the transmission request from the managed computer 1100 or may be the timing of a change in the storage configuration information table 4003 or the storage performance information table 4004.

[0039] The performance management program (performance management processor) 3001 of the managed computer 1100 then executes the performance analysis process (step S11006). The improved configuration determination program (improved configuration determination processor) 3004 executes an improved configuration determination process (step S11007) and transmits the performance management screen to the computer 1000 used by the administrator (step S11008). The CPU executes codes equivalent to the programs to realize the programs as the processors in parentheses.

[0040] <Performance Analysis Process>

Figure 12 is a flow chart for explaining details of the performance analysis process (step S11006).

[0041] The performance management program (performance management processor) 3001 first calculates the average of IO patterns (read rate: ratio of READ_IOPS in the total of READ_IOPS and WRITE_IOPS, read hit rate: HIT) of the storage performance information table 4004 (step 12001). The average herein denotes an average within a predetermined period (for example, a period from NOW to previous K months in Figure 16) of each RG (RAID group), and the averages of all data (RGs) in Figure 10 included within the predetermined period are calculated.

[0042] The performance limit calculation program (performance limit calculation processor) 3002 then executes a performance trend calculating process of components (step S12002). The details of the process are shown in Figure 13. The "components" herein denote controllers, disks, and host paths (items listed in Figure 6).

[0043] The performance limit calculation program (performance limit calculation processor) 3002 further calculates the performance limit of storage (step S12003). Specifically,

the minimum value of the performance limits of the components obtained in step S12002 is set as the performance limit of the storage 1400. Thus, the minimum value of the CONTROLLER 6003, the HDD 6004, and the HOST_PATH 6005 in the performance limit information table 3005 is set. This process is based on the notion that the performance limit of the component that would reach the limit first determines the performance limit of the storage. In this case, although the performance limit of the storage 1400 is calculated and acquired when necessary, the configuration is not limited to this. A performance limit corresponding to the configuration may be calculated and stored in a memory in advance to be read out and acquired therefrom.

[0044] Subsequently, the performance trend calculation program (performance trend calculation processor) 3003 executes a performance trend calculation process (step S12004). The details of the process are shown in Figure 14.

[0045] <Performance Limit Calculation Process of Components>

Figure 13 is a flow chart for explaining details of a performance limit calculation process of the components.

[0046] The performance limit calculation process program (performance limit calculation processor) 3002 reads out the storage configuration information table 4003 (see Figure 9) for each RAID group and selects a leading record (information in the leading row) (step S13001).

[0047] Next, the performance limit calculation process program (performance limit calculation processor) 3002 determines whether all records are processed (step S13002), ends the process when the determination result is true, and reads out the IO patterns of the target RAID groups calculated in step S12001 described above when the determination result is false (step S13003).

[0048] Subsequently, based on the combination of the target configuration (current configuration) and the target IO pattern, the performance limit calculation process program (performance limit calculation processor) 3002 acquires information of the performance limits of the disk controller, the disk, and the host path and updates the CONTROLLER 6003, the HDD 6004, and the HOST_PATH 6005 in the performance limit information table 3005 (step S13004). The information of the limitations is stored in advance in a table not shown, corresponding to the combination of the target configuration and the target IO pattern.

[0049] The performance limit calculation process program (performance limit calculation processor) 3002 further determines whether all records are processed (step S13005) and calculates the total of the performance limits of the disks when the determination result is true (step S13006). The performance limits of the RGs are added (in step S13006) to obtain the performance limit of the entire storage 1400, and whether the performance limits of the components of the last RG are obtained is checked in step

S13005 when there are a plurality of RGs in the storage 1400.

[0050] The performance limit calculation process program (performance limit calculation processor) 3002 also judges the performance limits (minimum values) of the components (controller and host path) (step S13007) and reads out the next record (step S13008).

[0051] <Performance Trend Calculation Process>

Figure 14 is a flow chart for explaining details of the performance trend calculation process.

[0052] The performance trend calculation program (performance trend calculation processor) 3003 reads out the storage performance information table 4004 (step S14001), uses, for example, Formulas 1 to 3 to calculate the performance trend that is an approximate straight line calculated from the storage performance information table 4004 as a performance trend, and updates the TREND 7001 in the performance trend information table 3006 (step S14002).

[0053] The performance trend calculation program (performance trend calculation processor) 3003 then calculates an optimistic performance trend from the calculated approximate straight line and a standard error (Formula 4) calculated from the storage performance information table 4004 and updates the OPTIMISTIC_TREND 7002 in the performance trend information table 3006 (step S14003). The performance trend calculation program (performance trend calculation processor) 3003 further calculates a pessimistic performance trend from the calculated approximate straight line and the standard error calculated from the storage performance information table 4004 and updates the PESSIMISTIC_TREND 7003 in the performance trend information table 3006 (step S14004).

[0054] The performance trend calculation program (performance trend calculation processor) 3003 then displays the obtained performance trend (approximate straight line), optimistic performance trend, pessimistic performance trend, and the performance limit (threshold) of the storage 1400 obtained in step S12003 on a performance graph 16004 (Figure 16) (step S14005). As a result, the administrator can recognize the period when the storage 1400 reaches the limit (limit reaching predicted time).

[0055] The formula used in the performance trend calculation process is as follows. In this case, the performance trend is, for example, a linear function, IOPS is y , and date and time is t .

[Math.1]

$$y=at+b \dots (\text{Formula 1})$$

To obtain the linear function using n pieces of data, a and b can be expressed as

follows.

[Math.2]

$$a = (n \times \sum(y \times t) - (\sum y \times \sum t)) / (n \times \sum(t^2) - (\sum t)^2) \dots \text{(Formula 2)}$$

[Math.3]

$$b = (\sum y - a \times \sum t) / n \dots \text{(Formula 3)}$$

[0056] Although the linear function expressing the performance trend is calculated using the least squares method in the present embodiment, the method for calculating the approximate straight line is not limited to the least squares method.

[0057] The standard error se can be expressed by the following formula.

[Math.4]

$$se = \sqrt{(\sum(1-y)^2 \div (N-2))} \dots \text{(Formula 4)}$$

The optimistic performance trend can be expressed by the following formula.

[Math.5]

$$y = a \times t + b - se \dots \text{(Formula 5)}$$

The pessimistic performance trend can be expressed by the following formula.

[Math.6]

$$y = a \times t + b + se \dots \text{(Formula 6)}$$

[0058] In the present embodiment, "standard error x 1" is added/subtracted from the performance trend to calculate the optimistic performance trend/pessimistic performance trend in order to include about 65% of all IOPS between the optimistic performance trend and the pessimistic performance trend. However, "standard error x 2" may be added/subtracted from the performance trend to calculate the optimistic performance trend and the pessimistic performance trend to include about 95% of all IOPS. Furthermore, although the standard error is calculated by the above formulas in the present embodiment because the linear function expressing the performance trend is calculated using the least squares method, the calculation formulas of the standard error are not limited by these.

[0059] <Improved Configuration Determination Process>

Figure 15 is a flow chart for explaining details of the improved configuration determination process.

[0060] The improved configuration determination program (improved configuration determination processor) 3004 determines RAID groups to be extended (step S15001), reads out the extension plan, and selects the leading record (step S15002). In the first embodiment, all RAID groups are RAID groups to be extended. Since the target RAID groups are narrowed down in the following second embodiment, the narrowing-down

process is equivalent to the process of determining the RAID groups to be extended. The extension plan is prepared in advance, and for example, an extension plan stored in the HDD 1106 of the managed computer 1100 is read out and stored in the memory 1107.

- [0061] Next, the improved configuration determination program (improved configuration determination processor) 3004 determines whether all records (extension plans) are processed (step S15003), ends the process when the determination result is true, and shifts the process to step S15004 when the determination result is false.
- [0062] The improved configuration determination program (improved configuration determination processor) 3004 then calculates the performance limits of the extended storages and updates the LIMIT 8007 in the improved configuration information table 3007 (step S15004). Thus, a process similar to steps S12002 and S12003 is executed in step S15004, and the performance limits of the storages are acquired from, for example, the above described tables based on the combinations of the extended configurations and the IO patterns before extension (calculated in S12001).
- [0063] Subsequently, the improved configuration determination program (improved configuration determination processor) 3004 calculates the time required for the extension, updates the TIME 8008 in the improved configuration information table 3007 (step S15005), reads out the next record (extension plan) (step S15006), and similarly calculates the time required for extension in the next plan. In this way, the time required for the extensions in all extension plans is obtained.
- [0064] The time required for extension is constituted by the total time of the time for creating the RAID groups in the storages, the time for creating the LUs in the storages, and the time for moving (copying) the data from existing LUs to extended LUs. The time for creating the RAID groups and the LUs is determined by their capacities. Supposing that the amount of data movement is determined by the ratio between the storage capacity before extension and the extended storage capacity, the time for data movement can be obtained by dividing the amount of data movement by the movement speed (transfer speed). However, since the time required for the data movement is longer than (dominant) the creation time of the RAID groups or LUs, that time can be set as the time required for extension.
- [0065] The RAID groups to be extended are LUs in which IOPS in the storage performance information table 4004 are not exceeding the pessimistic performance trend. However, the administrator can also select the RAID groups to be extended.
- [0066] <Configuration of Performance Management Screen>
Figure 16 is a diagram of a configuration of a performance management screen 16000. The performance management screen 16000 is constituted by, for example: a text box 16001 for inputting target storages of performance management; a button

16002 for ordering the execution of the performance management process; an extension plan list 16003 for displaying extension plans denoting the result of the performance management process after the update button is pressed; and a performance graph 16004 showing the current performance trend and the performance limit, the performance limit after improvement, and the time required for the improvement.

[0067] As a result, the performance trend of storage and the performance limit excess time can be accurately predicted without being affected by the peak value, and an optimal improved configuration can be proposed. Therefore, proactive performance management can be realized. Thus, in the example shown in Figure 16, considering the extension plan using a pessimistic prediction result, it can be determined that it is preferable to select the PLAN #2 that can complete the extension before exceeding the threshold of the performance limit.

[0068] 2) Second Embodiment

In the second embodiment, the performance management program 3001 of the managed computer 1100 executes a process of narrowing down the records of the storage performance information table 4004 used in the performance analysis process. As the configuration of the storage system is the same as in the first embodiment, the description will be omitted. The points in the present embodiment different from the first embodiment will be mainly described.

[0069] <Internal Configuration of Memory>

Figure 17 is a diagram of a memory configuration of the computer 1000. Upon startup, the computer 1000 reads out, from the HDD 1006 for example, a data input/output program 2001 for inputting and outputting data to the storage 1400, a computer setting management program 2002 for managing setting information of the computer, a computer setting table 2003 that is setting information of the computer, and an operation information table 17001 that is an execution schedule setting for operating jobs on the computer and stores them in the memory 1007.

[0070] The configurations of the memories of the managed computer 1100 and the storage 1400 are the same as in the first embodiment (Figures 2 and 3), and the description will be omitted.

[0071] <Contents of Read Out Tables>

The programs and the tables other than the operation table are the same as in the first embodiment, and the description will be omitted. The operation table 17001 is information used for narrowing down the RAID groups to be processed.

[0072] Figure 18 is a diagram of a configuration of the operation information table 17001. The operation information table 17001 comprises START_TIME 18001 denoting start date and time of jobs, PASSED_TIME 18002 denoting maximum execution time of the jobs, STORAGE 18003 that is an identifier of storages for the jobs, LU 18004 that

is an identifier of LUs for the jobs, and JOB 18005 denoting job names to be executed.

[0073] <Performance Management Process>

Figure 19 is a flow chart for explaining details of a performance management process in the second embodiment.

[0074] The administrator transmits a display request of the performance management screen (see Figure 16: PLAN, performance graph, and the like are not displayed at the initial stage) to the managed computer 1100 (step S11001). The performance management program (performance management processor) 3001 of the managed computer 1100 that has received the request transmits the performance management screen 2100 to the computer used by the administrator (step S11002). The computer used by the administrator can be any computer, such as the managed computer 1100 or the computer 1000, or may be completely different computers.

[0075] The computer used by the administrator then transmits a request of performance limit calculation and improved configuration display to the managed computer 1100 (step S11003). The performance management program (performance management processor) 3001 of the managed computer 1100 that has received the request transmits a transmission request of the storage configuration information table 4003 and the storage performance information table 4004 to the storage 1400 (step S11004).

[0076] The storage information management program (storage information management processor) 4002 of the storage 1400 that has received the transmission request transmits the storage configuration information table 4003 and the storage performance information table 4004 to the managed computer 1100 (step S11005). The timing of the transmission may be the timing of reception of the transmission request from the managed computer 1100 or may be the timing of a change in the storage configuration information table 4003 or the storage performance information table 4004.

[0077] The performance management program (performance management processor) 3001 of the managed computer 1100 also transmits a transmission request of the operation information table 17001 to the computer 1000 (step S19001). The computer setting management program (computer setting management processor) 2002 of the computer 1000 that has received the request transmits the operation information table 17001 to the managed computer 1100 (step S19002).

[0078] The performance management program (performance management processor) 3001 of the managed computer 1100 then executes a narrowing-down process of the storage performance information table 4004 (step S19003). Subsequently, the managed computer 1100 executes the performance analysis process (step S11006) and the improved configuration determination process (step S11007) and transmits the performance management screen to the computer used by the administrator (step S11008). The details of the performance analysis process and the improved configuration deter-

mination process are the same as in the first embodiment.

[0079] <Narrowing-Down Process (Determination Process of Extension Targets)>

Figure 20 is a flow chart for explaining details of the narrowing-down process.

The performance management program (performance management processor) 3001 of the managed computer 1100 reads out the operation information table 17001 and selects the leading record (step S20001). The performance management program (performance management processor) 3001 determines whether all records (all information in the operation table) are processed (step S20002)

[0080] When the determination result is false, the performance management program (performance management processor) 3001 calculates the time from START_TIME 18001 to START_TIME 18001 + PASSED_TIME 18002 in the operation information table 17001 as a job execution period (step S20003).

[0081] Next, the performance management program (performance management processor) 3001 determines whether the job execution period obtained in step S20003 is registered in a job execution time zone, i.e. in a table in the memory 1107 of the managed computer 1100 (concept: see Figure 21) (step S20004), and then reads out the next record when the determination result is true (step S20006) and registers the job execution period to the job execution time zone when the determination result is false (step S20005). A period with overlapping jobs is checked by the process of step S20005.

[0082] When the determination result of step S20002 is true, the performance management program (performance management processor) 3001 narrows down the records on the condition that the TIME in the storage performance information table 4004 is included in the job execution time zone (step S20007) and ends the process.

[0083] Figure 21 is a conceptual diagram of a table for showing definition of the job execution period and the job execution time zone shown in the narrowing-down process and for determining the job execution time zone. Reference numeral 21001 denotes periods of job execution. The time zones are one-day cycle and the time zones are specified only by the time because execution time zones of jobs started up at a specific time are illustrated in the present embodiment. However, the cycle may be one week or the cycle may be a one-month cycle. Furthermore, the time zones may be specified by the date and time without determining the cycle.

[0084] Job execution periods 21002 denote the times of execution of the jobs, and job execution time zones 21003 denote assemblies of execution periods 21002.

[0085] The administrator may be able to specify with which of the job execution time zones or time zones other than the job execution time zones the performance analysis will be performed. The administrator can also narrow down the records of the operation information table 17001 used for calculating the job execution time zones by the day or job name that can be calculated from the START_TIME 18001. The performance

management program (performance management processor) 3001 uses the records, in which the TIME 10001 of the storage performance information table 4004 is included in the job execution time zones, in the performance analysis process and the improved configuration determination process.

[0086] <Other Examples of Narrowing-Down Process>

Another narrowing-down method includes a method of narrowing down with used LUs determined from the computer setting table 2003. The used LUs denote an identifier of LUs used by the computer and denote a logical addition of LUs in which the JOB 5004 is not blank (= a certain job is using the LU) in the records of the computer setting table 2003.

[0087] The administrator may be able to specify whether to perform the performance analysis with the used LUs or with LUs other than the used LUs. The administrator can also narrow down the records of the operation information table 17001 used for determining the used LUs by the job names. The performance management program 3001 uses the records, in which the LUs in the storage performance information table 4004 are included in the used LUs, in the performance analysis process and the improved configuration determination process.

[0088] Another narrowing-down method of the storage performance information table 4004 includes a method of narrowing down with pessimistic performance trend excess LUs calculated from the storage performance information table 4004. The pessimistic performance trend excess LUs denote an identifier of LUs exceeding the pessimistic performance trend and denote a logical addition of LUs in which the total value of the READ_IOPS 10005 and the WRITE_IOPS 10006 of the same record is larger when a value calculated by the formula of the PESSIMISTIC_TREND 7003 in the performance trend information table 3006, with the TIME 10001 in the records in the storage performance information table 4004 being t , and a total value of the READ_IOPS 10005 and the WRITE_IOPS 10006 of the same record are compared.

[0089] The administrator may be able to specify whether to perform the performance analysis with the LUs exceeding the pessimistic performance trend or with LUs other than the LUs exceeding the pessimistic performance trend. The performance analysis process needs to be performed before the narrowing-down method is performed.

[0090] The performance limit after extension varies depending on the number of extended HDDs in the extended configuration. Therefore, it is effective to change the number of extended HDDs depending on the job type of the computer. As in the narrowing-down process described above, the performance trend of each job type can be calculated by executing the performance analysis process for each JOB 5004 in the computer setting table 2003. In this case, when the slopes of the linear functions of the performance trends of the jobs are different, the number of extended HDDs can be changed by the

rate of the slopes.

[0091] 3) Conclusion

In the present invention, the trend of workload (performance trend) is approximated with a function (for example, approximation with a linear function) based on time-series data of the amount of IO within a past determined period going back from now, and the standard deviation in the predetermined period is further obtained from the amount of IO. The function obtained by approximate function +/- standard deviation and the performance limit of the storage 1400 are then compared, and the time of reaching the limit is predicted and informed to the administrator. In this way, the accurate time of limit can be predicted without being affected by the instantaneous peak value of the workload, because the approximate function and the standard deviation are used and the prediction is based on the performance trend obtained first.

[0092] The prediction of the limit excess based on the performance trend and the standard deviation is performed every time the administrator requests. Therefore, the trend of workload in each case can be reflected on the prediction, and the administrator can recognize an appropriate limit excess time.

[0093] Furthermore, not only the total value of the performance limits of the disk devices of each RAID group in the storage 1400, but also the performance limit of the controller 1403 of each RAID group and the performance limit of the host path 1401 from the computer 1000 to the storage 1400 of each RAID group are taken into consideration, and the minimum value among them is set as the performance limit of the storage 1400. As a result, the limit excess time is predicted based not only on the capacity of HDD but also on the performances of all components related to reading and writing so that the limit excess time can be accurately predicted.

[0094] Furthermore, the extension time in the case where the configuration of a prepared extension plan of HDD is implemented is calculated in the present invention, the extension time including the time required for moving the data to the extended disk devices. The extension plan and the extension time are displayed on the display screen along with the comparison information. As a result, an optimal improved configuration can be presented based on a further accurate limit excess time, and proactive performance management can be realized.

[0095] Furthermore, the performance trend is calculated from the amount of IO in the time zone in which the job is executed or in another time zone, or the targets are narrowed down to the RAID groups including storage areas with the amount of IO exceeding the value of the performance trend plus the standard error to calculate the performance trends of the targets. In this way, the records are narrowed down before the performance analysis. As a result, the speed of the prediction computation can be increased, and the prediction accuracy of the trend analysis and the performance limit

excess time can be improved. The operation cost can also be reduced.

[0096] The present invention can also be realized by a program code of software realizing the functions of the embodiments. In that case, a storage medium recording the program code is provided to a system or a device, and the system or the device of a computer (or CPU or MPU) reads out the program code stored in the storage medium. In that case, the program code read out from the storage medium realizes the functions of the embodiments, and the program code and the storage medium storing the program code constitute the present invention. Examples of the storage medium for supplying the program code include a flexible disk, a CD-ROM, a DVD-ROM, a hard disk, an optical disk, a magneto-optical disk, a CD-R, a magnetic tape, a nonvolatile memory card, and a ROM.

[0097] An OS (operating system) or the like operated on the computer may execute part or all of the actual processes based on an instruction of the program code, and the processes may realize the functions of the embodiments. Furthermore, after the program code read out from the storage medium is written in the memory of the computer, the CPU or the like of the computer may execute part or all of the actual processes based on an instruction of the program code, and the processes may realize the functions of the embodiments.

[0098] The program code of the software realizing the functions of the embodiments may be distributed through a network to store the program code in storage means such as a hard disk and a memory of a system or a device or in a storage medium such as a CD-RW and a CD-R, and the computer (or CPU or MPU) of the system or the device may read out and execute the program code stored in the storage means or the storage medium before use.

Claims

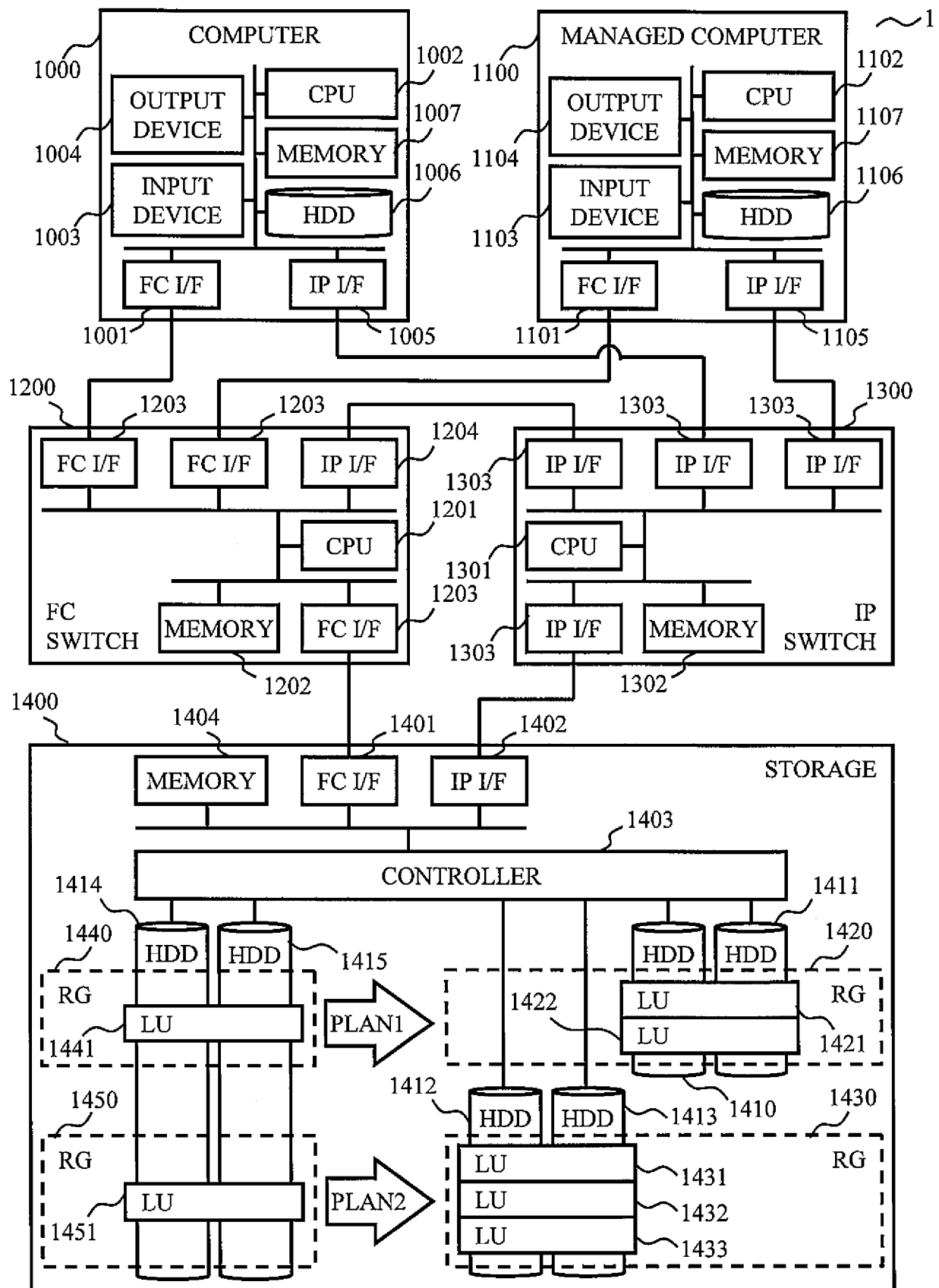
- [1] A storage system (1) comprising:
a storage (1400); and
a performance managing unit (1100) that manages processing performance of the storage, wherein
the performance managing unit obtains an approximate function of an amount of IO based on the transition of the amount of IO from a computer (1000) to the storage within a predetermined period to set the approximate function as a performance trend (7001) of the storage, further calculates a standard deviation of the amount of IO, and outputs the performance trend and comparison information between the standard deviation and the performance limit of the storage to an output unit (16004).
- [2] The storage system according to claim 1, wherein
the performance managing unit is a managed computer (1100),
the storage (1400), the managed computer (1100), and the computer (1000) are connected through a network,
the approximate function is an approximate straight line obtained from the transition of the amount of IO from an input of instruction to a past predetermined period,
the managed computer (1100) further comprising:
a performance limit calculation processor (3002) that sets a minimum value among a total value of performance limits of disk devices (1410 to 1413) in each RAID group in the storage (1400), a performance limit of a controller (1403) of the storage in each RAID group, and a performance limit of a bandwidth (host path (1401)) from the computer to the storage in each RAID group as a performance limit of the storage; and
an improved configuration determination processor (3004) that reads out a prepared extension plan of the disk device in the storage (1400), calculates an extension time in the case where the configuration of the extension plan is implemented, the extension time including a time required for moving the data of the extended disk devices, and outputs (16003) the comparison information as well as the extension plan and the extension time to the output unit.
- [3] The storage system according to claim 1, wherein
the performance managing unit (1100) calculates the performance trend and the standard deviation every time a computation instruction is inputted and outputs information of comparison with the performance limit of the storage to the output unit.

- [4] The storage system according to claim 1, wherein the performance managing unit (1100) sets a minimum value among a total value of performance limits of the disk devices (1410 to 1413) in each RAID group in the storage (1400), a performance limit of the controller (1403) of the storage (1400) in each RAID group, and a performance limit of the bandwidth (host path 1401) from the computer (1000) to the storage (1400) in each RAID group as a performance limit of the storage (1400).
- [5] The storage system according to claim 4, wherein the performance managing unit (1100) calculates an IO pattern for each RAID group, the IO pattern including an average of read/write rate and an average of read hit rate, and refers to a performance limit table storing the performance limits of the disk devices (1410 to 1413), the controller (1403), and the host path corresponding to combinations of configurations of the IO patterns and the RAID groups to acquire the performance limits of the disk devices, the controller, and the host path from configurations of the calculated IO patterns and RAID groups to be processed.
- [6] The storage system according to claim 1, wherein the performance managing unit (1100) reads out a prepared extension plan of the disk devices in the storage (1400), calculates an extension time in the case where the configuration of the extension plan is implemented, the extension time including the time required for moving data to the extended disk devices, and outputs the comparison information as well as the extension plan and the extension time (16003) to the output unit.
- [7] The storage system according to claim 1, wherein the performance managing unit (1100) calculates a time zone (21003), in which a job is executed, from execution schedule information (17001) of the job in the computer or calculates the performance trend from an amount of IO of another time zone to calculate the approximate function.
- [8] The storage system according to claim 6, wherein the performance managing unit (1100) takes the sum of a plurality of job execution periods to specify the time zone in which the job is executed.
- [9] The storage system according to claim 1, wherein the performance managing unit (1100) narrows down the targets to RAID groups including storage areas having an amount of IO exceeding the value of the performance trend plus the standard error and calculates the performance trend of the targets.
- [10] A program for controlling a performance managed computer (1100) that manages processing performance of a storage (1400) in a storage system (1), the

program causing the performance managed computer (1100) to realize:
a function of obtaining an approximate function of an amount of IO based on the transition of the amount of IO from a computer (1000) to the storage (1400) within a predetermined period to set the approximate function as a performance trend (7001) of the storage (1400); a function of calculating a standard deviation of the amount of IO; and a function of outputting the performance trend and comparison information between the standard deviation and the performance limit of the storage (1400) to an output unit (16004).

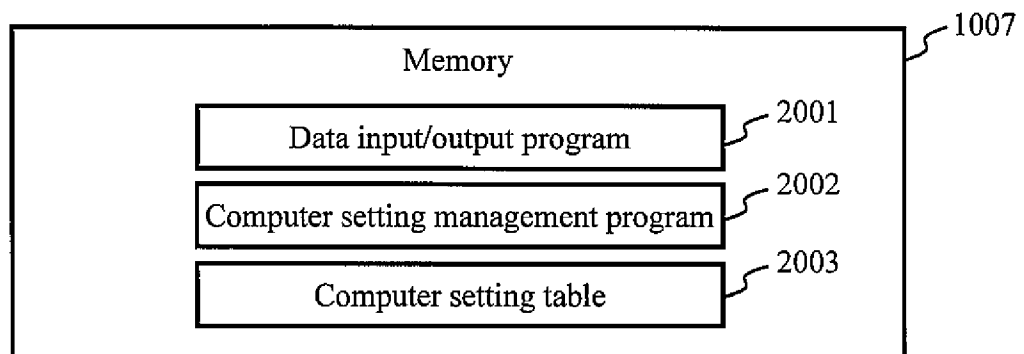
[Fig. 1]

FIG. 1



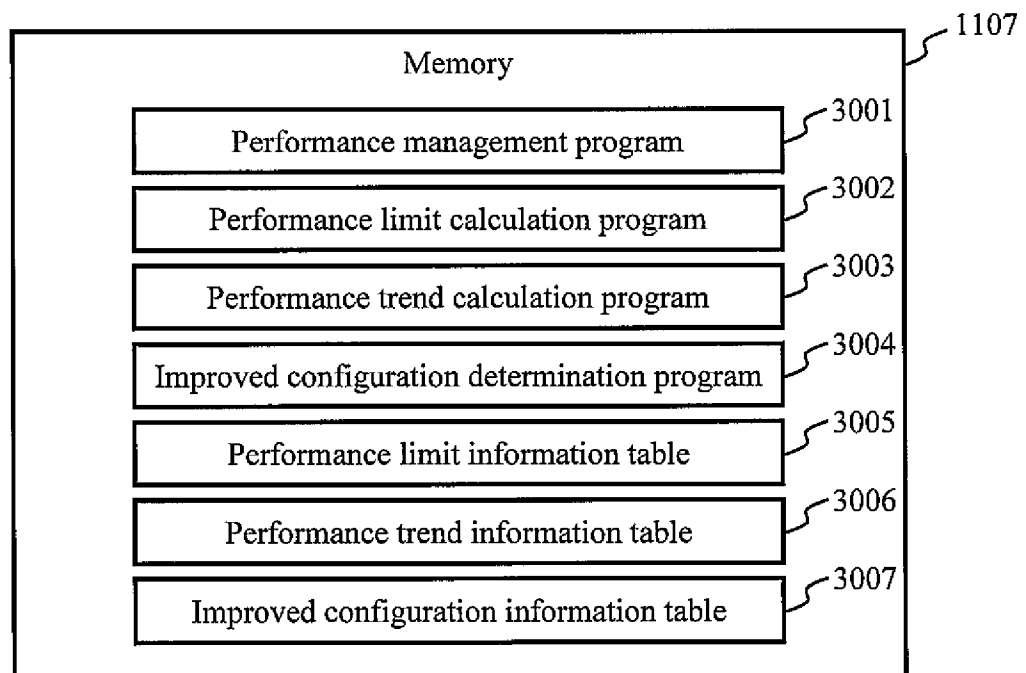
[Fig. 2]

FIG. 2



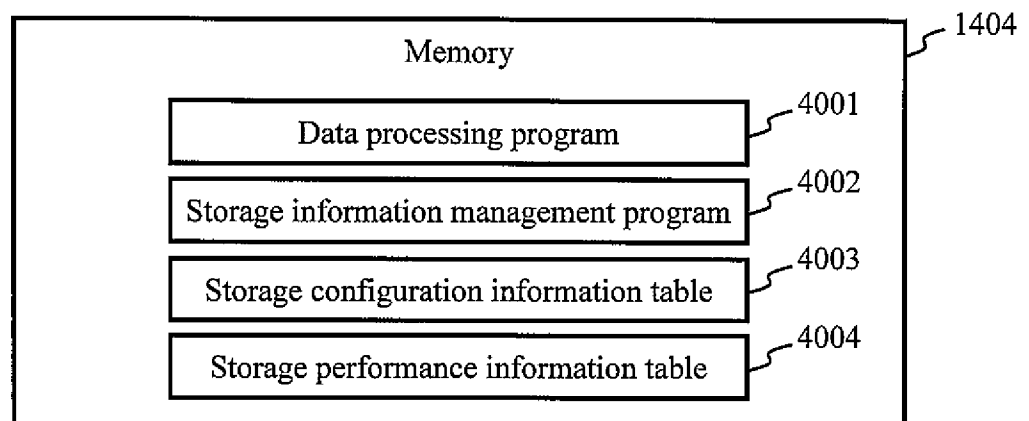
[Fig. 3]

FIG. 3



[Fig. 4]

FIG. 4



[Fig. 5]

FIG. 5

5001 COMPUTER	5002 LU	5003 TARGET IF	5004 JOB	2003
1000	1421	1401	BACKUP	
1000	1422	1401	BACKUP	
1000	1431	1401	ONLINE JOB	
1000	1432	1401	ONLINE JOB	
1000	1432	1401	ONLINE JOB	

[Fig. 6]

FIG. 6

6001 STORAGE	6002 RG	6003 CONTROLLER	6004 HDD	6005 HOST PATH	3005
1400	1420	260	300	400	
1400	1430	280	280	410	

[Fig. 7]

FIG. 7

7001 TREND	7002 OPTIMISTIC TREND	7003 PESSIMISTIC TREND	3006
$y=2t+10$	$y=2t+10-100$	$y=2t+10+100$	

[Fig. 8]

FIG. 8

8001 STORAGE	8002 HDD	8003 RG	8004 RAID LEVEL	8005 LU	8006 SIZE	8007 LIMIT	8008 TIME	3007
1400	1414&1415	1440	RAID1	1441	300GB			
1400	1414&1415	1450	RAID5	1451	300GB			

[Fig. 9]

FIG. 9

9001 STORAGE	9002 RG	9003 RAID LEVEL	9004 LU	9005 SIZE	4003
1400	1420	RAID1	1421	200GB	
1400	1420	RAID1	1422	200GB	
1400	1430	RAID5	1431	200GB	
1400	1430	RAID5	1432	300GB	
1400	1430	RAID5	1433	300GB	

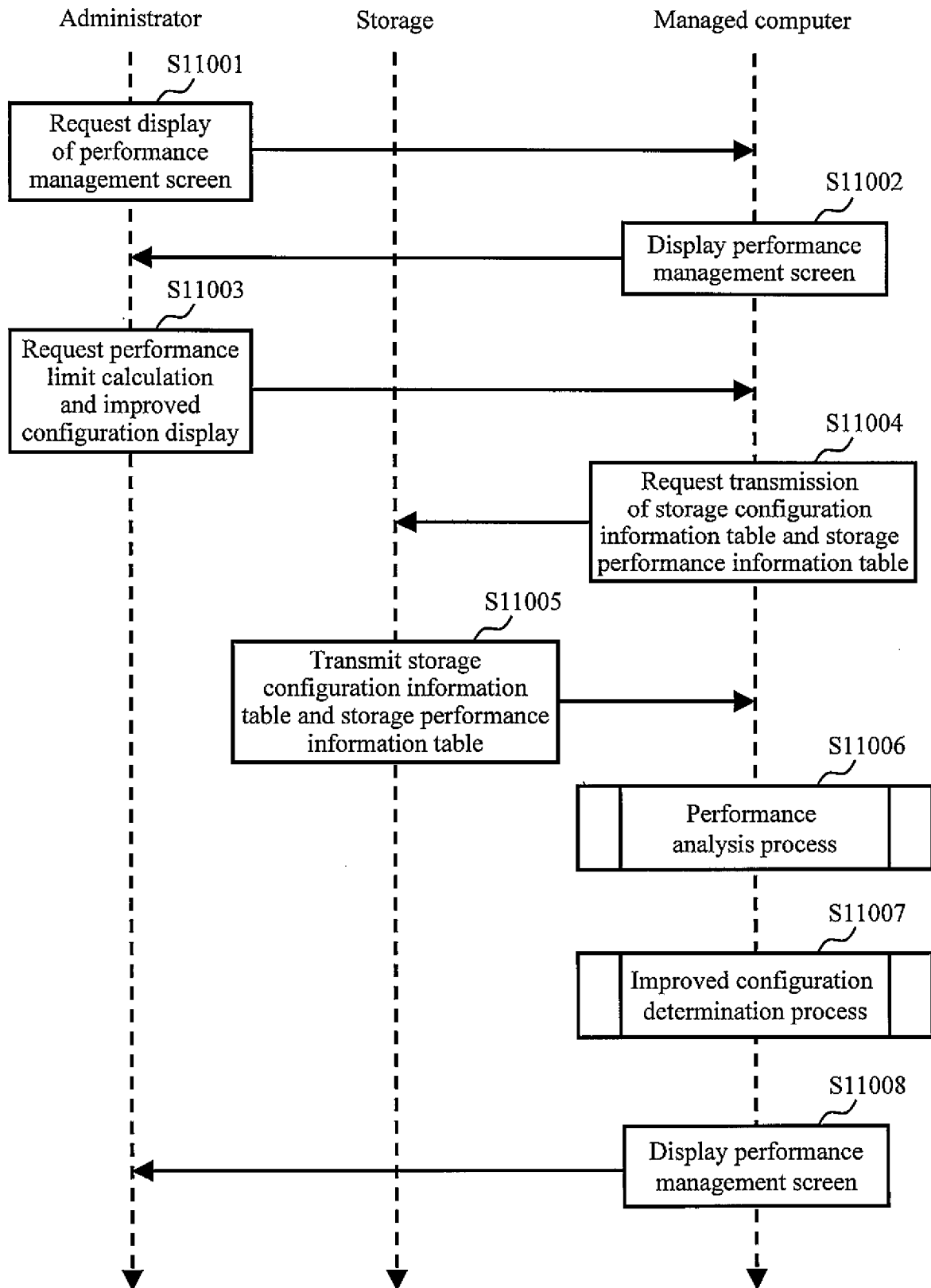
[Fig. 10]

FIG. 10

10001 TIME	10002 STORAGE	10003 RG	10004 LU	10005 READ IOPS	10006 WRITE IOPS	10007 HIT	4004
2008/11/20-10:00	1400	1420	1421	100	200	80	
2008/11/20-10:10	1400	1420	1421	200	300	70	
:	:	:	:	:	:	:	
2008/11/20-13:00	1400	1430	1431	10	50	10	
2008/11/20-13:10	1400	1430	1431	10	50	10	
:	:	:	:	:	:	:	
2008/11/20-13:00	1400	1430	1433	10	50	10	
2008/11/20-13:10	1400	1430	1433	10	50	10	
:	:	:	:	:	:	:	

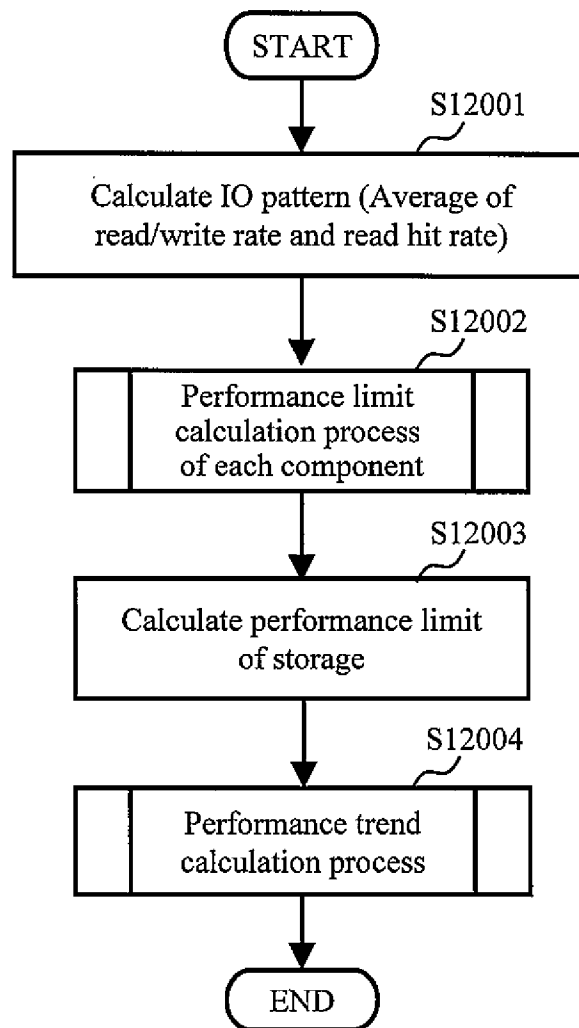
[Fig. 11]

FIG. 11

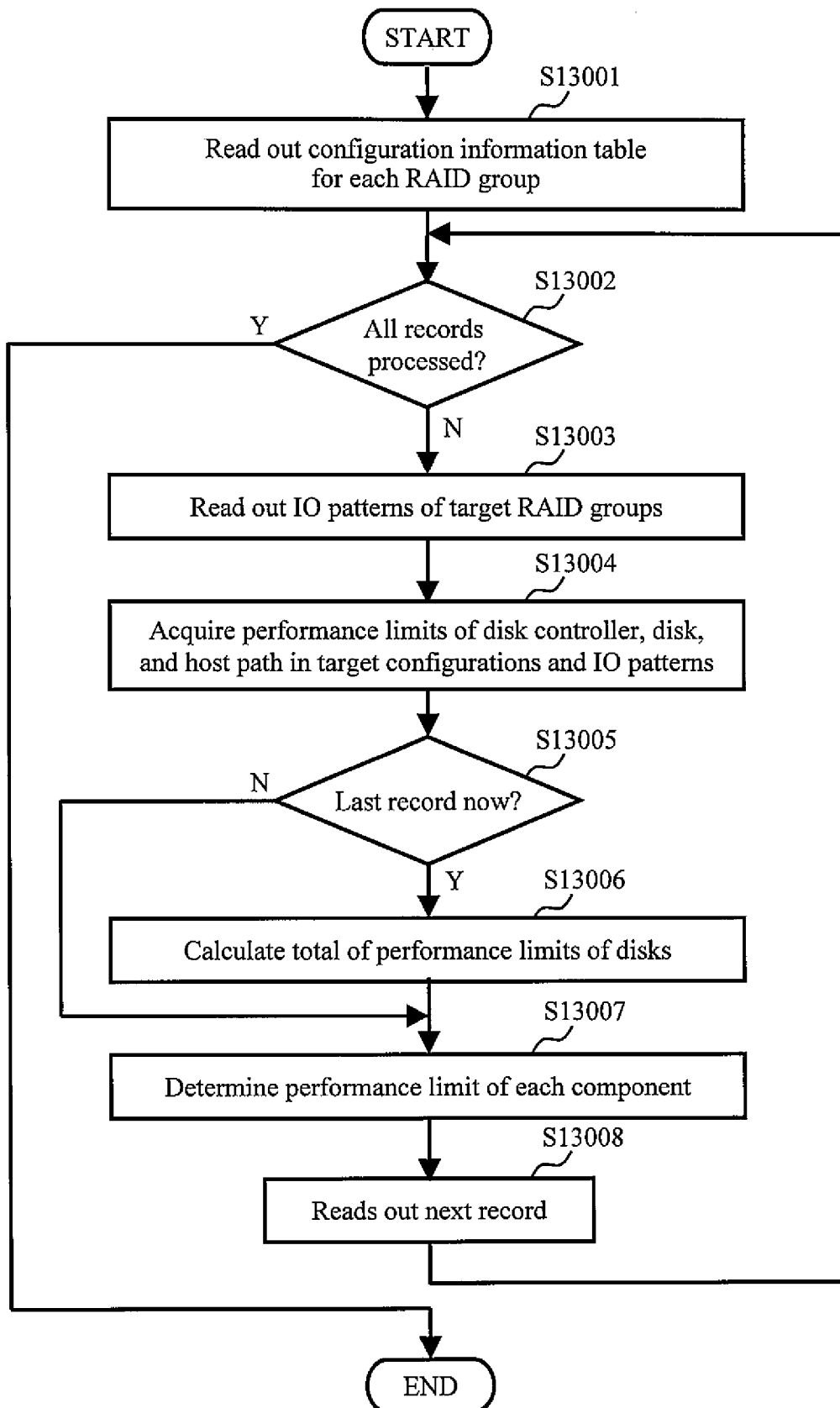


[Fig. 12]

FIG. 12

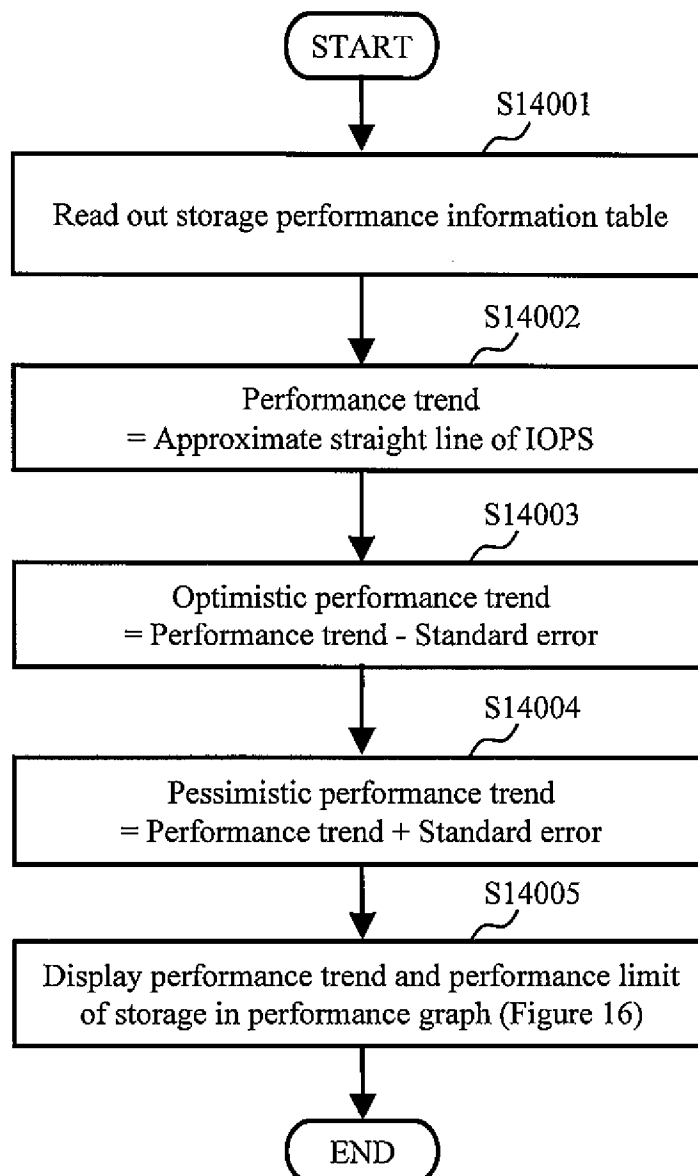


[Fig. 13]
FIG. 13



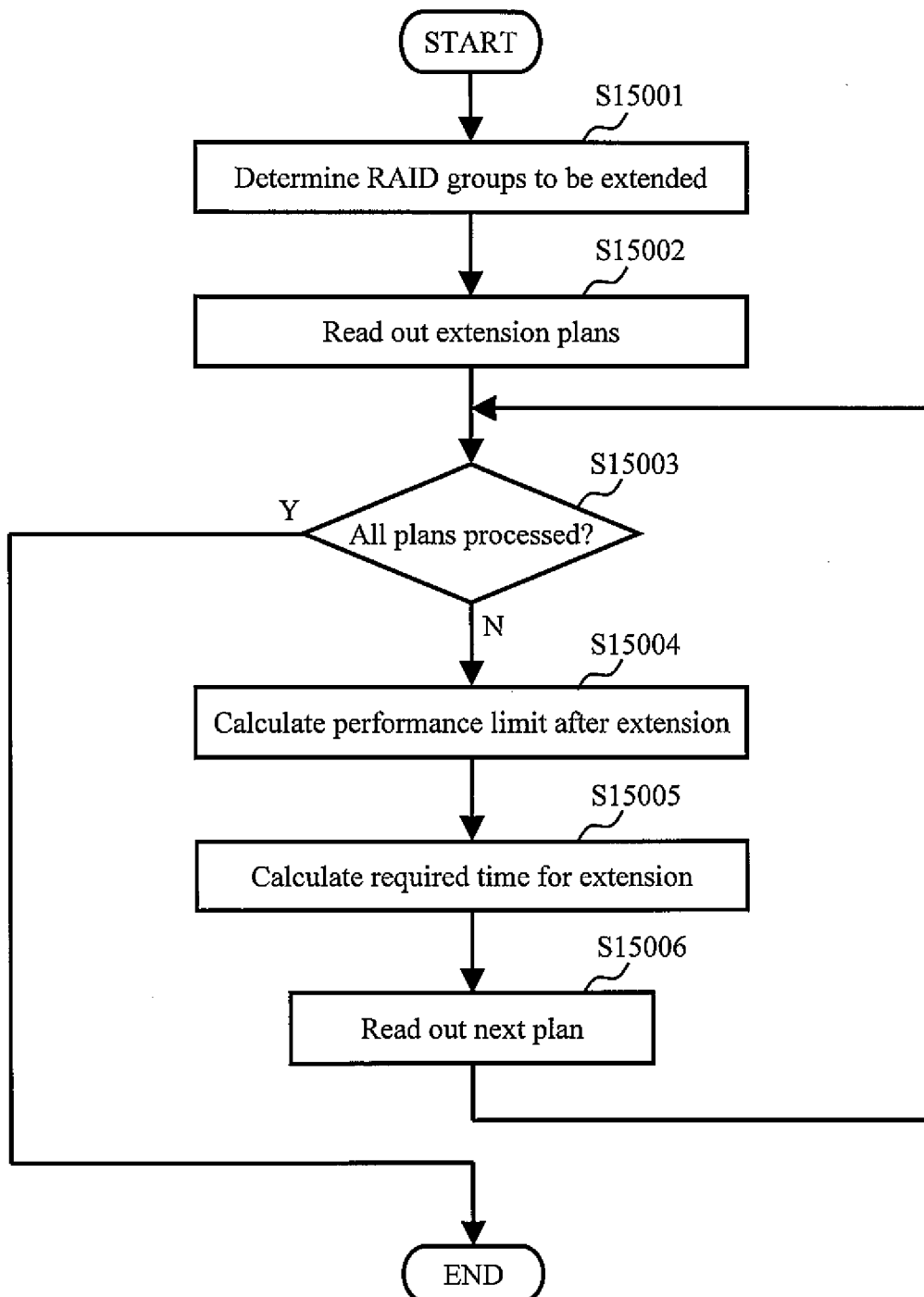
[Fig. 14]

FIG. 14



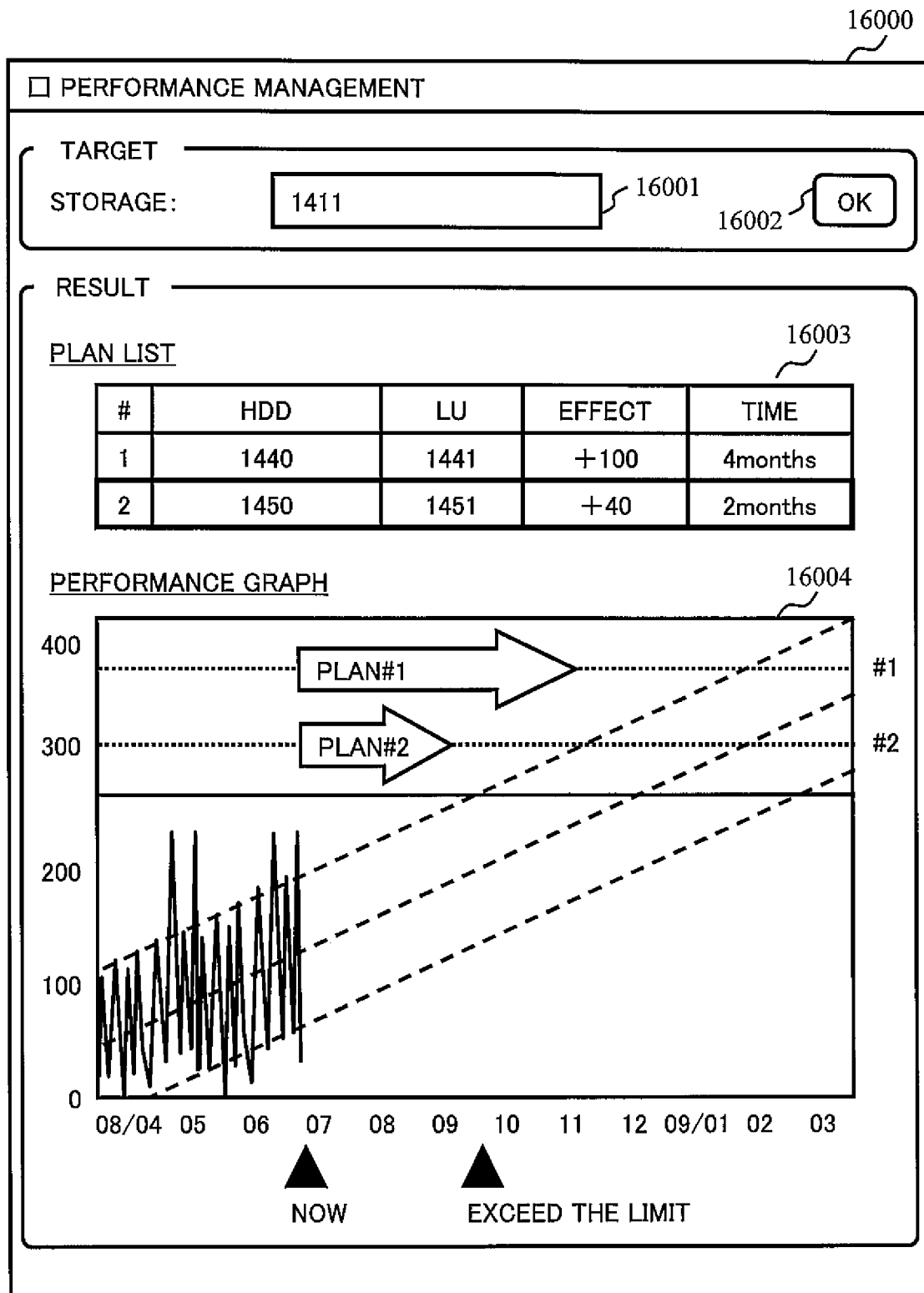
[Fig. 15]

FIG. 15



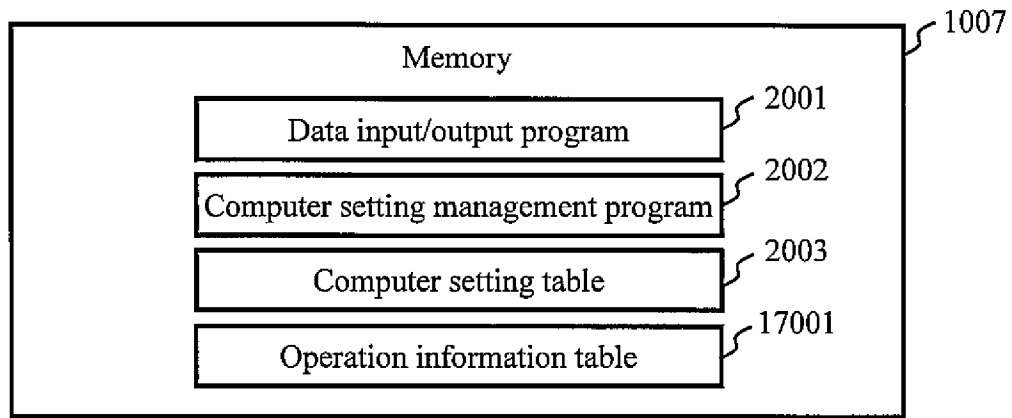
[Fig. 16]

FIG. 16



[Fig. 17]

FIG. 17



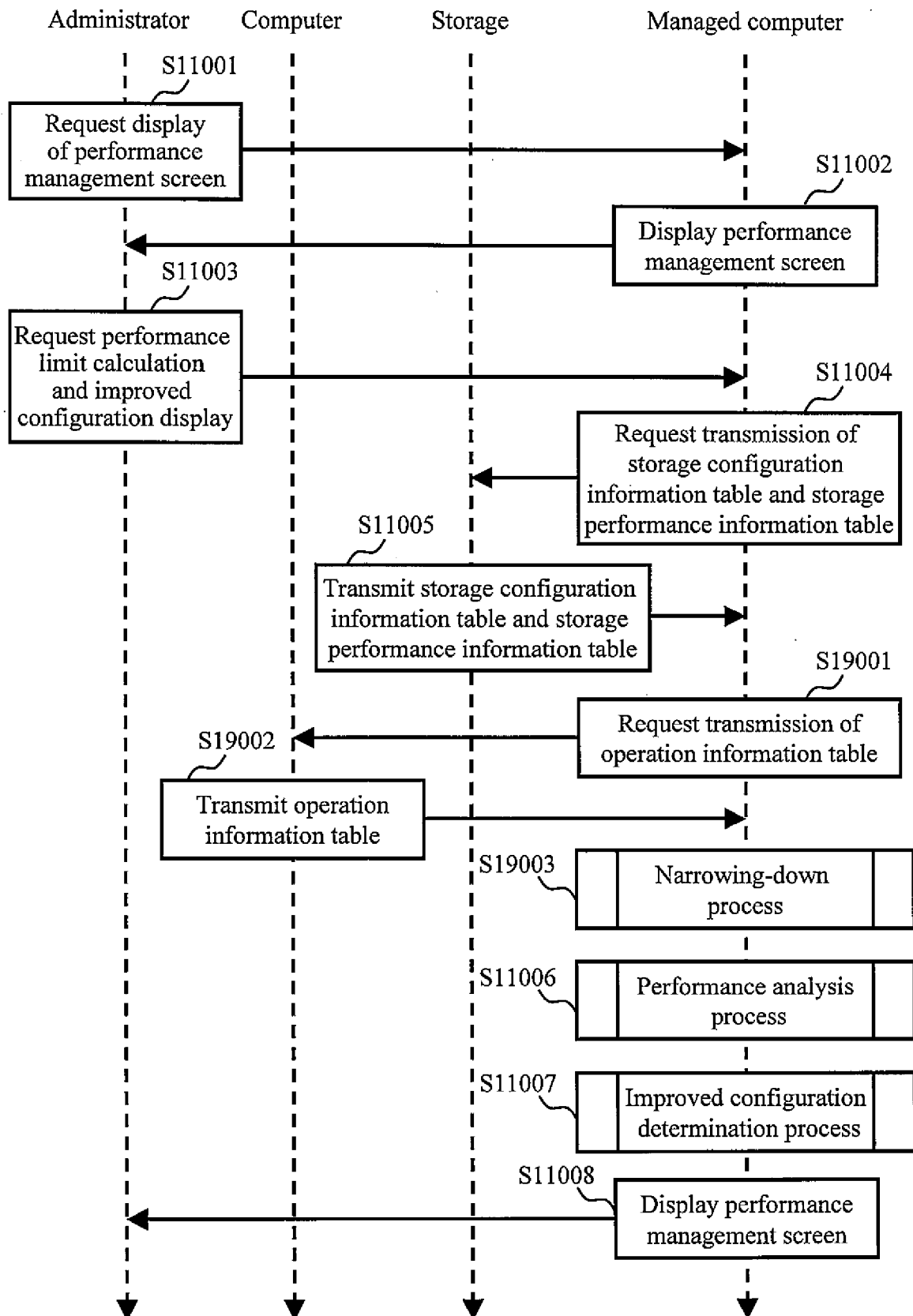
[Fig. 18]

FIG. 18

18001 START TIME	18002 PASSED TIME	18003 STORAGE	18004 LU	18005 JOB	17001
2007/04/01-10:00	180min	1400	1411	Backup1	
2008/04/01-10:00	180min	1400	1412	Backup2	

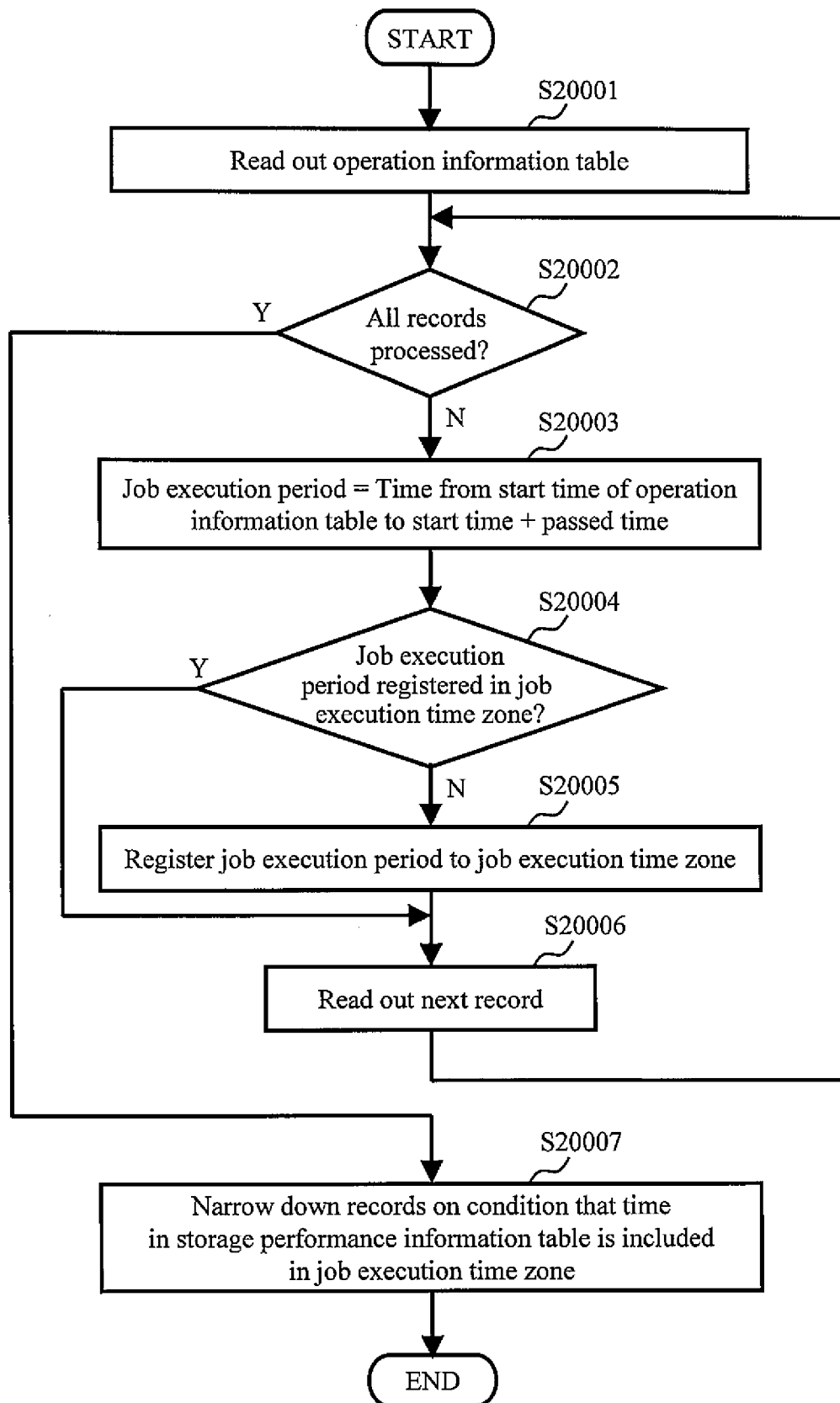
[Fig. 19]

FIG. 19



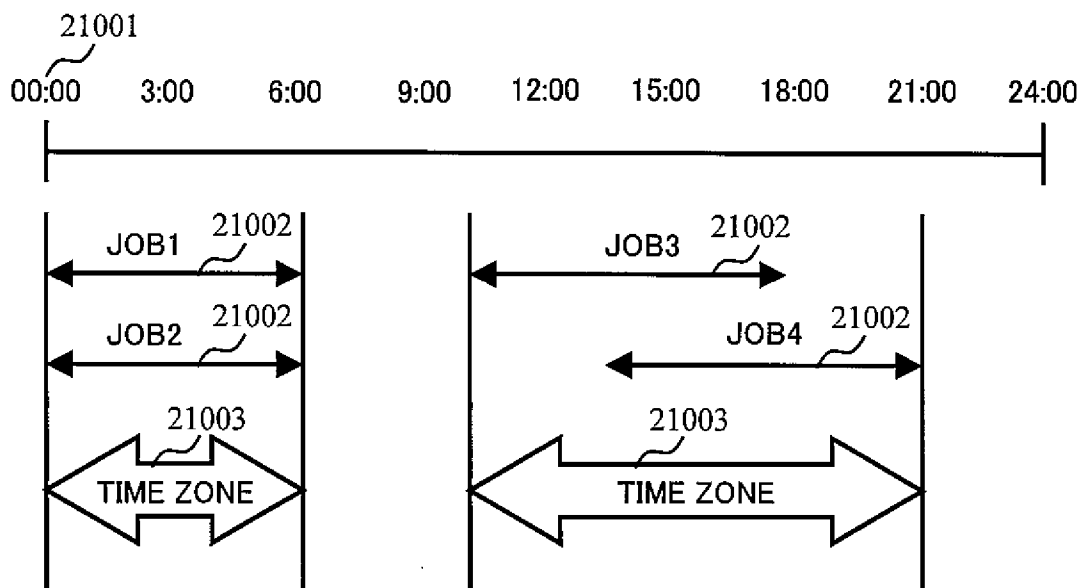
[Fig. 20]

FIG. 20



[Fig. 21]

FIG. 21



INTERNATIONAL SEARCH REPORT

International application No
PCT/JP2009/000504

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F3/06 G06F11/34

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 739 558 A (IBM [US]) 3 January 2007 (2007-01-03) paragraphs [0008], [0039] figure 2	1-10
X	US 2003/093619 A1 (SUGINO SHOJI [JP] ET AL) 15 May 2003 (2003-05-15) paragraphs [0035], [0058], [0073], [0074], [0086], [0087] figures 1,6,8	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

25 May 2009

Date of mailing of the international search report

02/06/2009

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

De Ceulaer, Bart

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/JP2009/000504

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 1739558	A	03-01-2007	NONE	
US 2003093619	A1	15-05-2003	JP 2003216348 A	31-07-2003