High-Speed Disk Device for the SX-9

FUJIMORI Hideaki, TAKIYANAGI Masumi, TANAKA Yoji

Abstract

The recent explosive dissemination of broadband has been increasing the volume of data in e-mails and movie contents at a rate of 160% per year and an assortment of data is now being exchanged via the Internet. As a result of this trend, the topics attracting special attention with regard to the types of storage used in IT systems are; the sudden increase in the data volume, complications related to data management and the important effects accompanying suspensions of operations. This paper discusses the storage device solutions that are being offered by NEC to confront these issues. These are the "iStorage D8" featuring particularly high scalability, manageability and availability and the "iStorage D1/D3,"which features an especially high cost efficiency, a simple introduction process and a space-saving design.

Keywords

storage, disc array device, RAID, cache control, unconstrained extensions
non-stop expansion, power consumption function (MAID), distributed access, backup

1. Introduction

In recent years the explosive dissemination of broadband communications has brought about a sharp rise in data exchange via the Internet. As a result, the volume of data handled in e-mails and movie content has been increasing at a rate of 160% per year, thus making it a matter of urgency for corporate IT systems to cope with the pressures of the new environment. In particular, the issues attracting special attention with regard to the storages used in supercomputer systems are as follows. Those related to sudden increases in data volumes, the complications associated with data management and the important effects accompanying suspension of operations. This paper will discuss the "iStorage D-Series" products, which are our solutions to such storage-related problems.

2. Customer Problems

The first of the problems for customers introducing storages in their systems is that concerning the measures to be taken to deal with unpredictable increases in data volumes. When a compact storage product is introduced in order to reduce initial costs but the data volume exceeds the maximum capacity of the compact storage, the purchase of a higher capacity product would be deemed to be necessary and the system would have to be stopped to allow data migration. In contrast, if a product with a large storage capacity is introduced initially, then the capital investment would be high.

The second issue occurs when several systems are built that are arbitrarily based on individual optimizations. In this case, the administration and management costs are increased due to the increased diversification of storages and the storage reconfigurations that accompany business modifications may be very complicated. For example, an increase in the number of operating storages makes their management more complicated. In addition, storage integrations using a large storage aimed at TOC reduction requires system stoppages in order to allow data migration as well as meticulous design work.

The third problem is the extension of operating hours following the globalization of businesses and both intra- and inter-corporate linkages of systems. Since the storages are always accessed by a large number of servers, a failure in the storage system often makes stoppages unavoidable. What is more, a system stoppage due to such a cause presents the risk of deteriorating the social credibility of the enterprise. As a result, the requirement for a non stoppage enabled performance is becoming particularly noticeable among supercomputer systems.

3. Storage Solution Proposed by NEC

As solutions from NEC for the issues outlined in Section 2

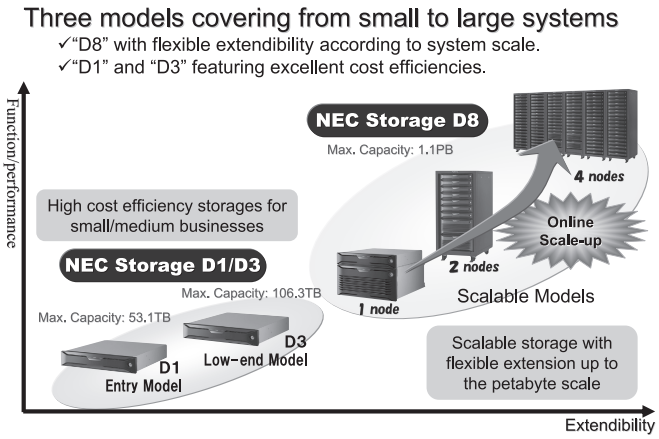

Fig. 1 Product lineup.

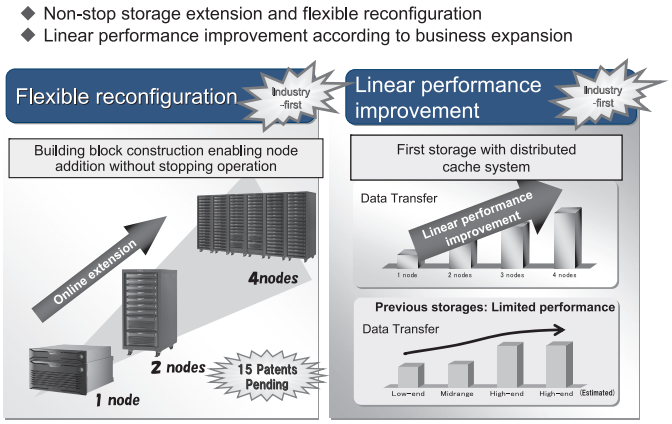

Fig. 2 Feature 1: Scalability.

above, we have developed the "iStorage D-Series" of next-generation scalable storage products (**Fig. 1**). The product lines including the iStorage D8 scalable storage capable of flexible extension at the petabyte scale and the iStorage D1/D3 high cost efficiency storages will respond to needs from the entry class to the high-end systems.

3.1 iStorage D8

iStorage D8 is a SAN (Storage Area Network) compatible storage product that has the following three big features.

(1) Scalability

Even when a system is started at a small scale its capacity can be increased up to the petabyte scale and its performance can be improved linearly according to the desired capacity increase (**Fig. 2**). There is a capability of increasing the number of nodes, which are the units used in the storage, from a minimum of 1 node to a maximum of 4 nodes without stopping operations. This allows the storage performance that has been limited with previous storage products to be improved linearly according to the capacity increase.

The technologies that have made this flexible reconfiguration based on node and module expansions possible are the industry-first building-block architecture and the distributed cache system (**Fig. 3**). With regard to the performance limitations of the previous design that were due to the concentration of accesses to a specific cache, the accesses are now distributed on a per-node basis. This strategy enables a linear improvement in the performance.

In addition, in order to improve the performance even further, an intelligent cache system is adopted to incorporate an MPU (Micro Processing Unit) in each cache module in order to execute complicated cache data controls by limiting them inside the cache module.

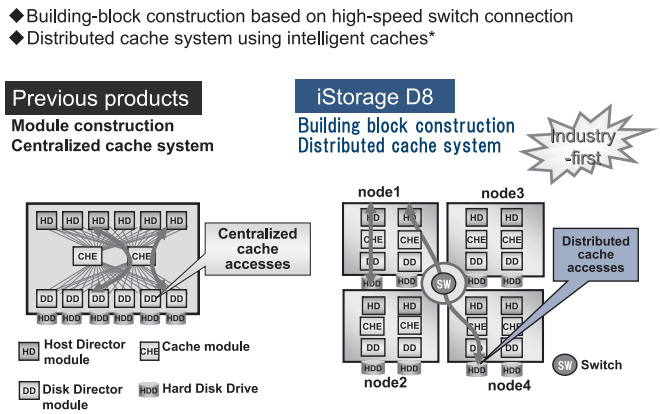

Fig. 3 Scalable technology.

(2) Manageability

iStorage D8 offers the iStorage Virtual Storage Partitioning function, which employs virtualization technology to allocate the resources required for jobs as desired.

This function pools the physical resources of the storage and forms virtual storages by allocating the resources in the pool according to the job type. Additions and modifications of resources in the virtual storages are thus facilitated without affecting the servers and applications. In addition, a user management function is provided for use in setting the administrator for each of the virtual storages. This function

restricts virtual storage access without the authority of the administrator, thereby preventing operation errors and illegal accesses and ensuring the secrecy of jobs (**Fig. 4**).
The resources can be allocated to the virtual storages in two ways according to the job type; either by using the performance emphasis method that allocates resources as proprietary resources or by the efficiency emphasis method that allocates them as shared resources (**Fig. 5**).

1) Performance Emphasis Type

The host ports, caches and disk pools are allocated to the virtual storages as proprietary resources. The allocation of proprietary resources can eliminate resource competitions between jobs and minimize the mutual effects on performances with other jobs.

2) Efficiency Emphasis Type

The resources are shared between jobs in order to optimize the utilization efficiency of the virtual storages. This method not only increases the disk pool capacity efficiency but also enables fine controls such as minimizing the size settings of caches, which results in important performance effects on a per-job basis.

3) Availability

The SPOF (Single Point of Failure) is thoroughly eliminated in order to meet the improved availability requirement of the SX-9 system. The possibility of system stoppages due to unexpected faults is minimized by means of multiplexing and modularization (**Fig. 6**).

◆ Unhindered allocation of resources for jobs using virtualization technology
◆ No need for server/application setting changes, even in the case of reconfiguration

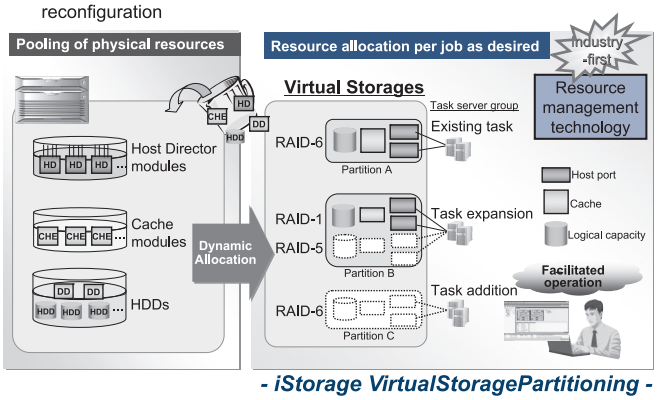

Fig. 4 Feature 2: Manageability.

◆ Elimination of SPOF* and minimization of system stoppages due to faults by means of multiplexing and modularization

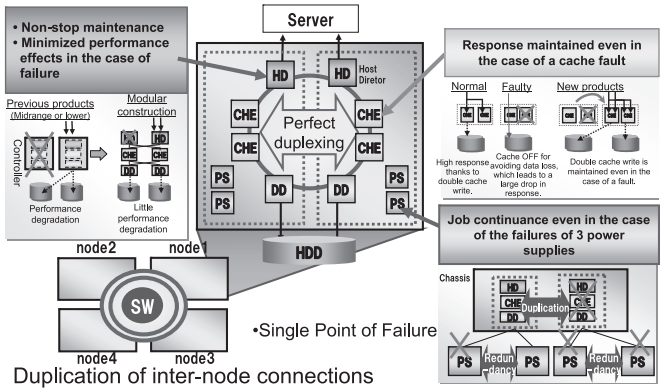

Fig. 6 Feature 3: Availability.

◆ Flexible physical resource allocation for optimum storage management according to the job type

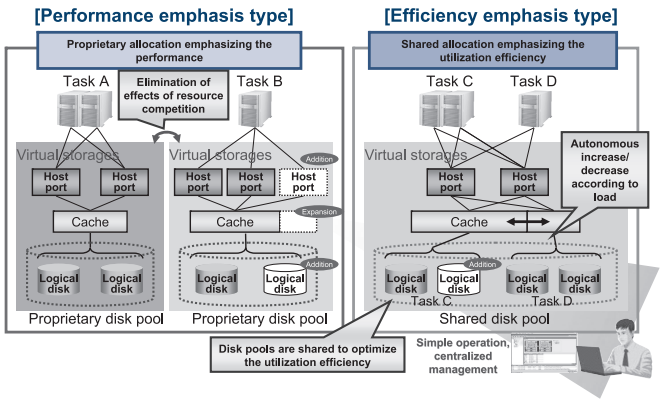

Fig. 5 Resource management technology.

♦ Job continuance even in the case of double HDD faults
♦ Industry-first triple mirror provided as standard

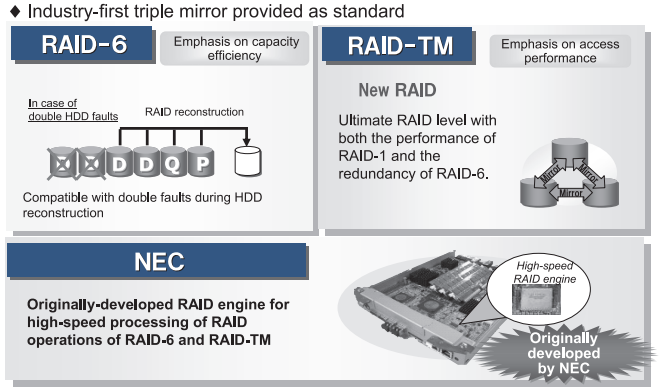

Fig. 7 RAID-6 and RAID-TM.

In addition to the RAID-6 (Redundant Array of Independent Disks) that makes job continuance possible even in the case of double faults in the HDD, the SX-9 also implements RAID TM (Triple Mirror), which has both the performance of RAID-1 as well as the redundancy of RAID-6 (**Fig. 7**). Whereas the previous products used the MPU in the controller for executing complicated RAID computations, the SX-9 incorporates our originally-developed LSI called the "high-speed RAID engine" in order to achieve high-speed processing.

3.2 iStorage D1/D3

iStorage D1 and iStorage D3 inherits the manageability and availability features of iStorage D8 at the same time as enabling high cost efficiency, simple introduction and space-saving design. The compact cabinet with a 2U size accommodates installation of a maximum of 12 HDDs and expansion of maximum 12 FC host interface ports. The maximum number of HDDs can be increased to 144 by connecting the DEs (Disk Enclosures) (iStorage D3). Multiplexing and redundancy are used in the same way as iStorage D8 and together with technologies such as RAID-6 and RAID-TM, achieve the high availability required of the SX-9 systems.

In addition, in order to meet the need for an introduction cost reduction, the simplified introduction function makes the introduction of storage easy for users without expert knowledge of storage issues. The initial setup wizard of the Webcam iStorage Manager navigates the series of introduction procedures from the installation of the storage to that of the software, construction of the storages and allocation of disks to servers. The storages also have a detection function, which automatically detects the servers to which each storage or disk introduced is to be allocated. These functions significantly reduce the time and labor required for the storage introduction, namely by reducing the labor by 1/3rd and the time to by 2/3rds compared to that required for the introduction of previous storage products (**Fig. 8**).

3.3 Energy Saving Function

As storage systems increase the scale, its power consumption and running costs are also increased. The "iStorage D-Series" uses the iStorage Storage Power Conserver that incorporates the MAID (Massive Arrays of Inactive Disks) in order to conserve the energy consumption of the system and make it friendly to the environment.

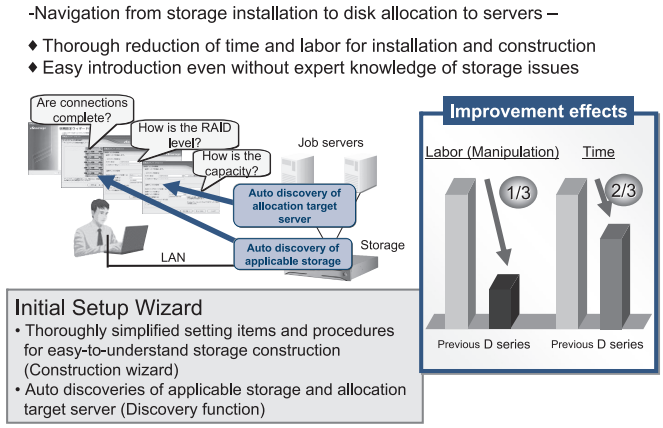

Fig. 8 Simplified introduction of storage.

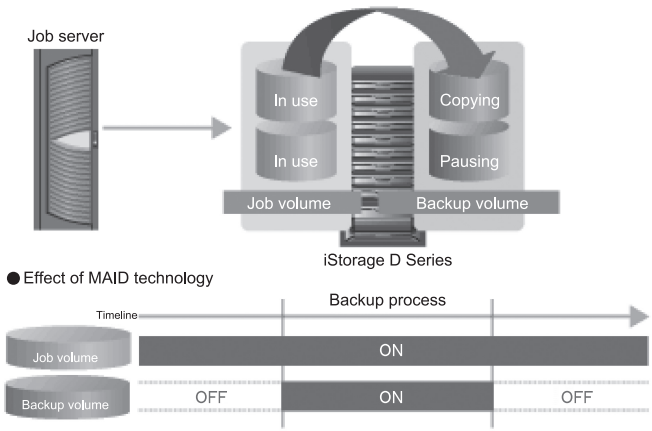

Fig. 9 Energy-saving operation with MAID technology.

The iStorage Storage Power Conserver controls the start and stop of the HDDs in the pool for a volume according to the operating status of the volume. It runs the HDDs that are not usually used, such as the backup HDDs, only when required and thus reduces the system power consumption by up to 30% (**Fig. 9**). It has acquired the NEC Eco Symbol as an environmentally friendly product for use in storage systems.

4. Conclusion

In the above, we discussed the "iStorage D-Series" solutions for the storage-related problems of customers by focusing on future system perspectives. In the future, we will enhance these products in a timely manner in line with the

changes in customer needs and market trends. We also intend to offer new storage products that are able to efficiently satisfy the needs of our customers.

Authors' Profiles

FUJIMORI Hideaki
Senior Manager,
System Storage Products Division,
1st Computers Operations Unit,
NEC Corporation

TAKIYANAGI Masumi
Manager,
System Storages Products Division,
1st Computers Operations Unit,
NEC Corporation

TANAKA Yoji
System Storage Products Division,
1st Computers Operations Unit,
NEC Corporation