Storage System: Management of Explosively Increasing Data in Mission-Critical Systems

By Toshiro NAKAJIMA,* Hirohiko KOIKE,* Hirokazu ISHIHARA,† Hiroyuki OHTANI‡, Hiroaki NAGANO‡ and Toshiyuki OHTSU‡

ABSTRACT The NEC Storage product family provides a storage platform to strategically utilize the explosively increasing data in enterprises. This paper describes the functions, which are based on NEC's platform technology, VALUMO, provided by the NEC Storage SAN (Storage Area Network) and NAS (Network Attached Storage) products, to reduce the TCO (Total Cost of Ownership) and improve availability and reliability.

KEYWORDS **SAN (Storage Area Network), NAS (Network Attached Storage), RAID (Redundant Array Independent Disk), Storage, Storage management**

1. INTRODUCTION

With the acceleration of broadband Internet and spread of the information technology in businesses, many enterprises are promoting global collaborations with partners in the world with rapidity. In this situation, it is obvious that the quantity of data accumulated in enterprises is increasing rapidly. Using ever-increasing data strategically is now one of the most important factors for strengthening the competitiveness of an enterprise.

The NEC Storage consists of a wide variety of storage products such as "S Series" SAN (Storage Area Network) products, "NS Series" and "NV Series" NAS (Network Attached Storage) products, and "T Series" tape library products.

The NEC Storage can operate storages effectively and correctly by virtualization for effective management of ever-increasing data, and by autonomous reaction against problems such as failures and performance degrading. The NEC Storage can also provide a disaster recovering function by the distribution of storages. In addition, the NEC Storage provides powerful storage solutions by strategic alliances with partners. The NEC Storage product family uses the VALUMO technology to reduce TCO (Total Cost of Ownership), improve system availability and reliability, and realize business continuity.

This paper describes the functions provided by the NEC Storage hardware and software of SAN and NAS series by using the VALUMO technology.

2. NEC STORAGE SAN PRODUCTS

The NEC Storage SAN products are configured with S series hardware and SAN software. This section describes the features of the products.

2.1 Hardware

The NEC Storage "S series" SAN products provide the highest level of performance, reliability, and availability in the world based on high-speed, high-scalability architecture including state-of-the-art high-speed X-bar technology, duplication and redundancy structures enabling high reliability, and the "Phoenix technology" (**Fig. 1**) which improves the data protection performance of RAID (Redundant Array Independent Disk) while providing the responses to the servers by utilizing HDD's autonomous recovery performance.

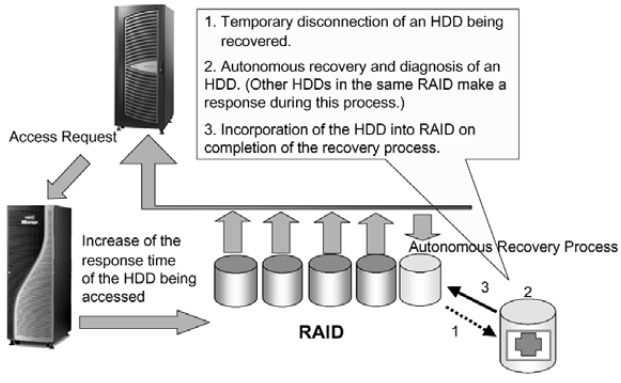

Fig. 1 Phoenix Technology.

*System File Products Division
†NEC System Technologies, Ltd.
‡1st Computers Software Division

2.1.1 Phoenix Technology

The HDD itself has advanced autonomous recovery capability against temporary malfunctions. However, a storage product cannot respond to an access command to the HDD being recovered. In order to provide high-speed responses to a server, a conventional storage product normally degrades and replaces the defected HDD without waiting for the completion of the recovery process, and rebuilds the data in the new HDD after the replacement.

This approach makes the system enter the non-redundant configuration until the data rebuild is complete. In recent years, the non-redundant duration has tended to become longer with the increase of HDD capacity. The longer non-redundant duration increases the risk of double faults and reduces the availability of disk array units. Using the phoenix technology, the NEC Storage disconnects an HDD being recovered autonomously for a moment without immediate degrading of the HDD. Then the NEC Storage incorporates the HDD into RAID again on completion of the HDD recovery process.

Disconnecting of HDD from RAID during the autonomous recovery processing can greatly reduce the non-redundant time and improve the data protection ability of the RAID.

2.1.2 RAID-6

NEC Storage S2800, the latest model of NEC Storage series, adopts RAID-6 (**Fig. 2**) to improve availability. RAID-6 is the RAID technology which actualizes super availability by two HDDs for keeping data redundancy while one HDD fails and enabling data access while even two HDDs fail, whereas RAID-1 and RAID-5 only enable data access while one HDD fails using one HDD for redundancy. The redundant data of RAID-6 consists of two parity data, "P" and "Q," which are generated by different logic from one data set. When two HDDs fail, the data will be restored by solving simultaneous equations made of the remaining data including P and Q. NEC Storage S2800 uses newly developed, exclusive LSI for the parity generation and complex calculation of data restoration.

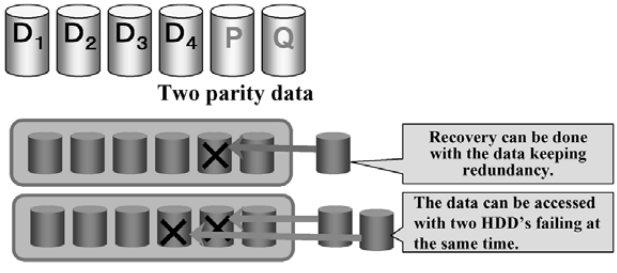

Fig. 2 RAID-6.

2.1.3 High-Speed, High-Reliability Architecture

The NEC Storage of S3000 and S4000 series, upper models in S series, adopts a high-speed, high-reliability architecture including the X-bar technology. **Figure 3** shows a conceptual diagram of the architecture.

Components are located separately in the left and right areas. The components in each area constitute their own cluster. These clusters are symmetric with each other and have redundant features. The HD (Host Director) is equipped with more than one microprocessor, and performs I/O control between the cache memory and the host computer and data transfer. The DD (Disk Director) is equipped with more than one microprocessor, and performs I/O control between the cache memory and HDDs and data transfer. The cache memory is used to cache an I/O data. Each of the HD and DD can access all cache memories and write data to the caches in both clusters concurrently. All of the HD, DD, and cache memories are connected with each other by using the X-bar technology. Using the X-bar technology, cache memory is connected to a node (HD or DD) in an end-to-end way to have a non-blocking structure in which access between a node and a cache is not affected by another node-to-cache access. The total throughput of the X-bar itself is up to 19.3GB/S (S4000 series) and thus the load fluctuation on performance is extremely low.

Furthermore, all components including power supply, fan, backup battery, and AC power receiving section are duplicated for high reliability.

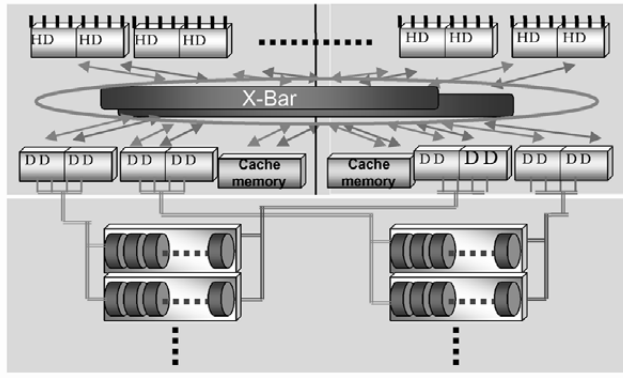

Fig. 3 High-speed, high-reliability architecture.

2.2 Software

The software products of NEC Storage S series consist of the Storage management product family of Intelligent Management Platform, which provides the integrated management of the system containing servers and storages, and the Storage platform product family of High Availability System Platform, which improves the availability of the system containing storage by controlling storage autonomously and virtually. These products constitute the VALUMO technology (**Fig. 4**).

The Storage management reduces TCO by providing the functions for realizing integrated management including the function of integrated management of storage resources and fault monitoring, the storage network management function, the function to support storage performance monitoring and analysis, and the collaboration function.

The Storage platform provides mission-critical infrastructures, which support 24-hour/365-day operation, including the data replication function for replicating storage volumes, the autonomous and virtual storage control function, and storage availability improvement function.

The following introduces NEC Storage Manager and NEC Storage PerformanceMonitor, which are typical storage management products, NEC Storage DynamicDataReplication and NEC Storage RemoteDataReplication, which are typical storage platform products, and NEC Storage CachePartitioning and NEC Storage DynamicSnapVolume, which is the latest priority product.

2.2.1 Storage Management Products

The NEC Storage Manager manages the storage configuration, monitors its operation status, and detects occurrences of faults in NEC Storage series disk array units. It allows an operator to centralize the management of multiple disk arrays of different models on a single client screen. A storage administrator can easily manage the configuration of several disk arrays and detect occurrences of faults. In addition, the collaboration of the integrated management software with NEC Storage Manager enables integrated management of servers and storages (**Fig. 5**).

The NEC Storage "PerformanceMonitor" monitors the information on the load and performance of NEC Storage series disk arrays. It provides the function to

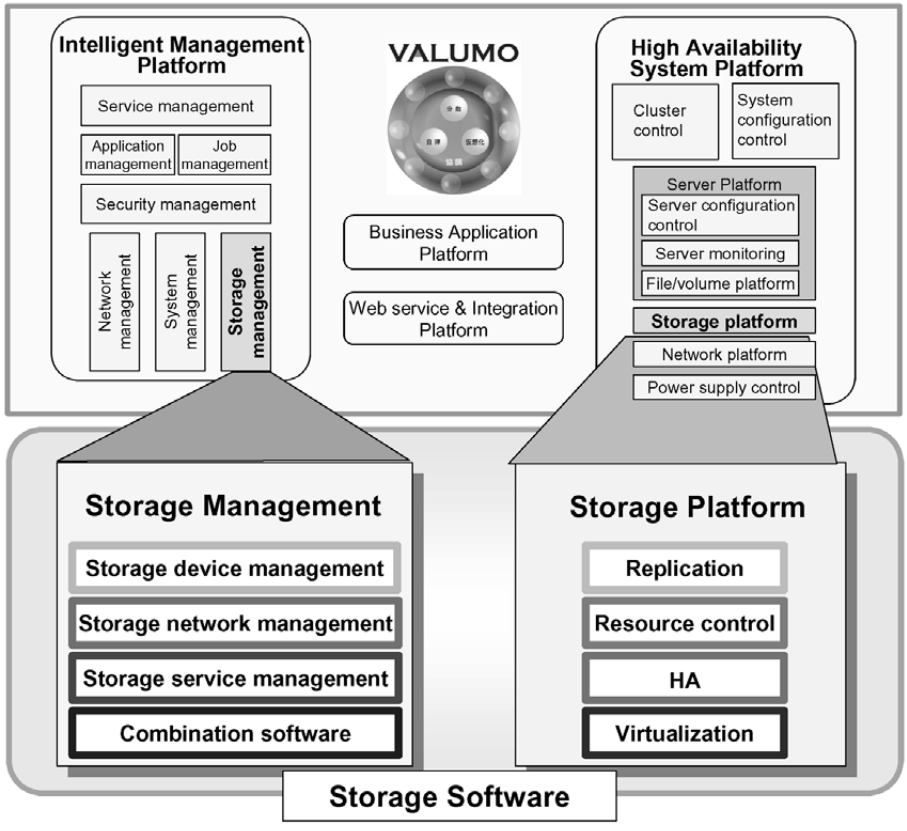

Fig. 4 S series software category.

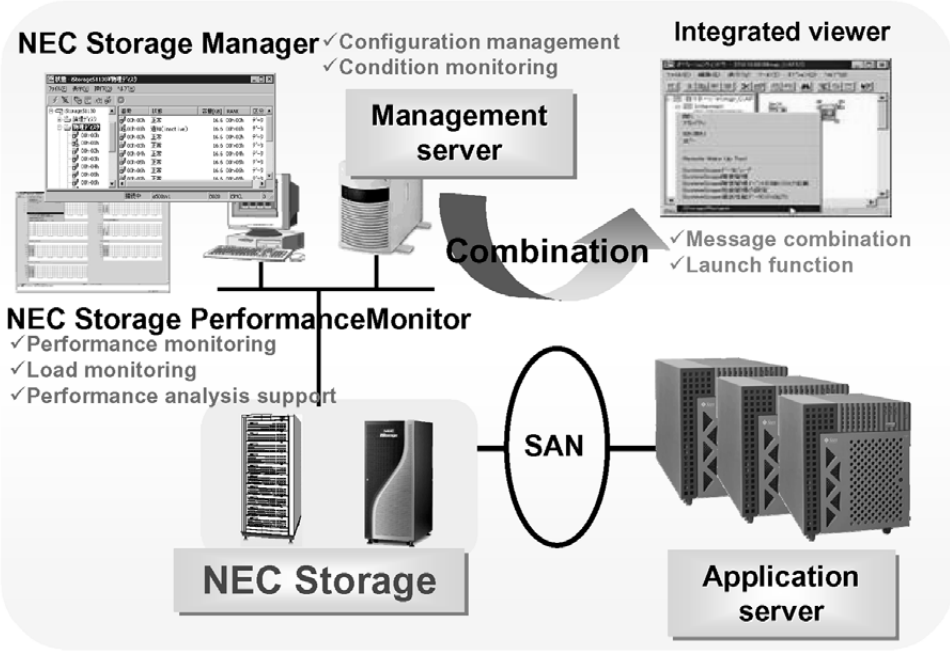

Fig. 5 Integrated management of server and storage.

visually display the performance information of the NEC Storage Manager on the client screen in real time, the function to monitor unexpected loads, and the function to support long-term performance analysis by accumulating performance information, and supports the performance tuning of storage units.

2.2.2 Storage Platform Products

The NEC Storage "DynamicDataReplication" provides the data replication function allowing replications of job volumes to be created in a disk array system with simple operations. The function can disconnect replicated volumes at any time and reduce processing time by performing backup processes in parallel with online jobs. Using replicated volumes, evaluation of applications in parallel with online jobs is possible (**Fig. 6**).

The NEC Storage "RemoteDataReplication" is a function to create replications of job volumes among several disk array systems. In a similar way to "DynamicDataReplication," it provides base functions including backup in parallel with job operations, data transfer to a different disk array system, and disaster recovery for providing against unexpected disasters by saving backups in a remote place.

NEC Storage DynamicSnapVolume (**Fig. 7**) is the function to make snapshot into the disk array unit. Whereas NEC Storage DynamicDataReplication is the function to make a complete copy of data volume, NEC Storage DynamicSnapVolume is the function to make a virtual copy by copying the difference of the data volume, and it is particularly effective for logical data damage such as program bug and human error. Since it is the differential copy, the copy requires only 20-30% of the original capacity, though it depends on the amount of data change. In that respect, NEC Storage DynamicSnapVolume is lower cost solution than NEC Storage DynamicDataReplication. NEC Storage DynamicSnapVolume can also manage the plural generation of copy, so that it is possible to keep the backup data of more recent recovery point by making frequent copies.

The NEC Storage "CachePartitioning" divides cache memory installed on a disk array into several segments (cache segments) to control the amount of cache memory occupied by each logical disk assigned to jobs.

In general, cache memory is shared by all jobs. Accordingly, the use ratio of cache memory fluctuates depending on the job operation status. It may bring an unstable I/O response performance. Accordingly, the NEC Storage "CachePartitioning" controls the cache memory occupation amount in jobs to eliminate the influences of minor jobs on major jobs. This can secure the I/O response performance and improve job stability. The reservation of cache memory allows a

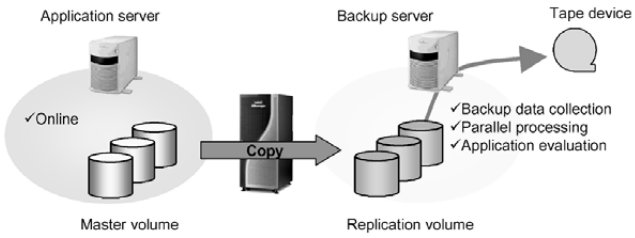

Fig. 6 Data replication.

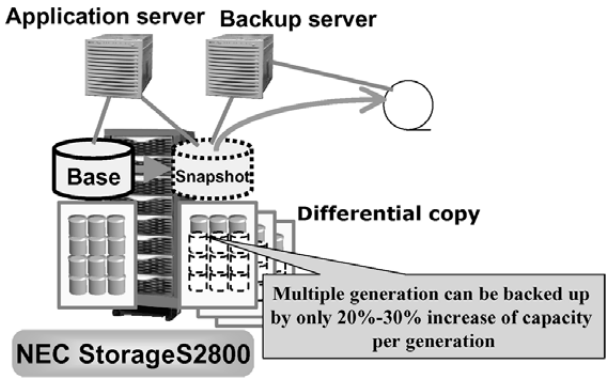

Fig. 7 DynamicSnapVolume.

new job to be added without any influence on jobs in progress. The creation of cache memory segments and the setting of occupation amount can be performed dynamically without job interruption.

3. NEC STORAGE NV SERIES

The NEC Storage NV series comprises high-end NAS products suitable for jobs requiring large capacity, high performance, and high availability. This section describes the features of the NV series products.

3.1 Hardware

The NV series products realize the highest level of performance and reliability in the class by state-of-the-art technologies including ultrahigh-speed TOE, large-scale NVRAMs, remote DMA cards, and high-speed HW RAID engine. **Figure 8** shows a block diagram of the NV series product.

(1) Ultrahigh-Speed TOE

The TOE is hardware performing the TCP/IP protocol processing, which conventionally has been performed by the CPU in a computer. The installation of the TOE allows the processing capacity of the CPU left by the off-loading to be used for NAS solutions.

(2) Large-Capacity NVRAMs (Non Volatile RAMs)

The main storage is configured with large-capacity non-volatile memory devices (NVRAMs) to return an I/O response against a write request as soon as the request is written to an NVRAM. Write data is accumulated in NVRAMs for batch writing. This allows response time to be reduced greatly in comparison with actual I/O operations to disks.

(3) High-Speed HW RAID Engine

Adopting 2Gbps realizes high performance. Furthermore, the addition of check frames can improve data reliability.

(4) High-Performance HDD

A high-reliability, dual port disk drive with 2Gbps fibre channel interface is used.

(5) Remote DMA

The HW engine is exclusively used for node-to-node communication in cluster configuration. The exclusive HW engine allows high-speed data transmission between nodes.

3.2 Software

3.2.1 Major Advantages

The NEC Storage NV series is controlled by software “NAS Basic Control,” which is configured with basic software “SC-LX (Storage Control Linux)” based on the open-source Linux technology and integrated management software “NEC Storage Manager IP” based on web-base GUI (Graphical User Interface).

The “NAS Basic Control” has the following functions:

(1) Network Control Function

The NAS Basic Control has the functions for improving usability and availability including control of the TCP/IP offload engine (TOE) originally developed by NEC and link aggregation.

(2) File Server Function

The NAS Basic Control includes a highly tuned file system and several network protocols such as NFS (Network File System), CIFS (Common Internet File System), and NDMP.

(3) Access Control Function

The NAS Basic Control provides UNIX-type file permission control and access control using Windows-type ACL (Access Control List). Share control

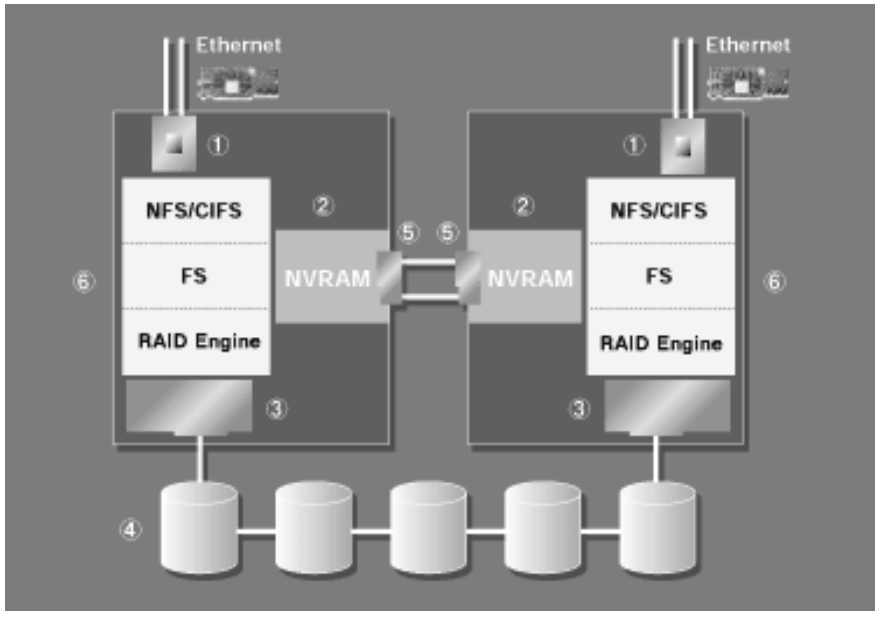

Fig. 8 NV Series block diagram.

between NFS and CIFS allows multiple clients of different types to share files by using the ID mapping function.

(4) High Availability Function

The NAS Basic Control connects the NEC Storage NV series in two-node configuration through high-speed communication bus to control clusters backed up by each other.

(5) Operation Management Function

The NAS Basic Control provides an easy operation environment for integrated management of NEC Storage NV series on web browsers.

Furthermore, the NEC Storage NV series provides various solution-based software products including those for replication, backup, and security (**Fig. 9**).

(6) Replication

The NAS Basic Control provides the replication function for replicating file system data in the same unit or another unit (MVD Sync supplied by Mountain View Data Inc.). Using the product enables the star-shape replications via IP network.

(7) Backup

Using the NDMP (Network Data Management Protocol) allows data backup and restoration to be controlled through networks by using backup software available for the NDMP. In addition, NetVault (backup software supplied by Bakbone Software Inc.) can be installed in the NEC Storage NV series to construct an autonomous backup system requiring no backup server.

(8) Security

The antivirus function for preventing computer virus infection (ServerProtect supplied by Trend Micro Incorporated.) can be installed in the NEC Storage NV series to prevent viruses penetrating into file systems without network load or use of an exclusive server for virus check.

Using these software products, the NEC Storage NV series can construct storage solutions appropriate to various business environments to contribute to the reduction of the TCO required for data management.

3.2.2 Software Base Technology

The NEC Storage NV series is configured by using various technologies for keeping the balance among high-performance, high-availability, operability, and adaptability required for the VALUMO technology.

On the basis of Linux, the software base SC-LX realizes high throughput by making full use of high-speed response and large-capacity resources through complete tuning by NAS appliance and implementation of real-time OS and expanded address technologies. In particular, the high-speed response is related to the performance in actual operation environments which may not appear in benchmark tests, and thus it is an important target in the development of the NEC Storage NV series.

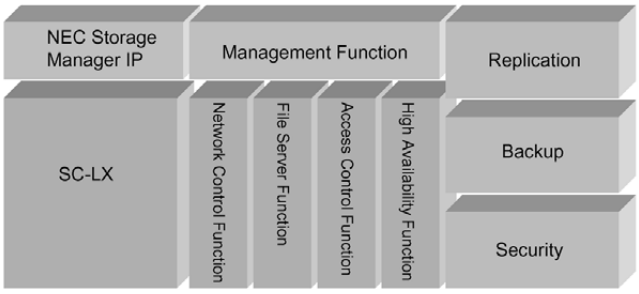

Fig. 9 Software structure of NV Series.

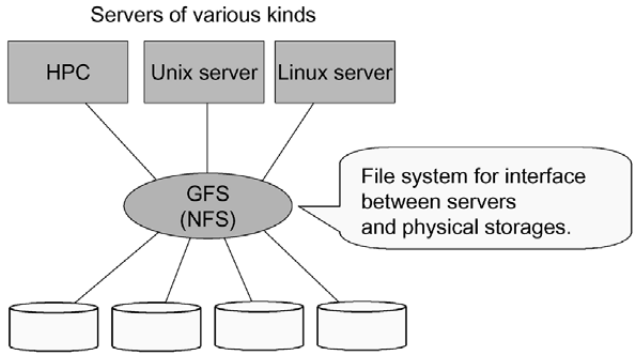

Fig. 10 Global File System.

For high availability, state-of-the-art speed-up technologies including the TOE and RAID engines are installed, and redundancy by multiple ports is also secured. In cluster configuration, consistent and bi-directional backup by sharing of information between nodes using the RDMA engine is provided to cope with high performance.

For operability, a snapshot of the file system level is adopted for backup and generation management to respond to flexible data handling in operation environments. The snapshot can be applied to various operations because it allows access to arbitrary files in images, which are unavailable for disk-level freezes. The integrated management base uses several technologies commonly used by PC servers such as NEC Express series servers to allow centralized monitoring together with other servers through NEC ESMPRO and other utilities.

For adaptability to the operation policy of a specific system, the NEC Storage NV series includes remote CLI permitting a variety of detailed controls in addition to GUI, and also adopts XML for describing in-system setting information in prospect of high-level system coordination control in the future. The use of Linux API in a UNIX (Linux)/Windows coexisting environment for which no standard operation is established currently can provide the user mapping and authentication between NFS and CIFS with a high degree of freedom to secure flexible adaptability.

The technologies described above are only some of the technologies applied to the NEC Storage NV series. However, it is expected that when installed, these technologies will play important roles as components of the system environment at which the VALUMO technology aims.

3.3 Future Enhancement Plan

(1) GFS (Global File System)

The GFS will be installed in the NEC Storage product family. The GFS is a high-performance and high-reliability file system for technological computing in a heterogeneous environment allowing servers with different OSs to link with each other and to accommodate storages other than NEC storage.

As shown in **Fig. 10**, the GFS as appears similar to an NFS which is widespread throughout the world from each server. It can handle as much data as 1 terabyte or more at high speed.

Owing to these advantages, the GFS is expected to play an important role in the technological computing field.

(2) NAS Switch

The NAS Switch virtualizes NAS devices to make more than one NAS appears to be only a single NAS for simplifying their operations. In addition, the NAS switch allows NAS devices to be added and data to be relocated during a job. This enables NAS devices to be shifted or upgraded without job interruption.

(3) NAS Gateway

The NAS Gateway can operate as a NAS by connecting to a SAN. This enables SANs and NAS devices to be managed totally for consolidation of storage including SANs and NAS devices.

4. CONCLUSION

This paper has described the VALUMO technology provided by the NEC Storage series. NEC will develop functions based on the VALUMO technology to provide customers with products whose performance, reliability, and availability have been improved further.

*UNIX is the registered trademark of The Open Group in the U.S.A. and other countries.

†Microsoft, Windows, Windows NT, MS-DOS, Active Directory are trademarks or registered trademarks of Microsoft Corporation in the U.S.A. and other countries.

‡Linux is the trademark or registered trademark of Mr. Linus Torvalds in the U.S.A. and other countries.

§NFS is the registered trademark of Sun Microsystems, Inc.

‖Other names of companies and products are trademarks or registered trademarks of each company.

Received April 13, 2004

* * * * * * * * * * * * * * * * *

Toshiro NAKAJIMA joined NEC in 1983, and is currently a manager of the System File Products Division.

Hirohiko KOIKE joined NEC in 1985, and is currently a manager of the System File Products Division.

Hirokazu ISHIHARA joined NEC in 1985, and is currently a manager of the Servers Platform Division of the NEC System Technologies.

Hiroyuki OHTANI joined NEC in 1995, and is currently an assistant manager of the 1st Computers Software Division.

Hiroaki NAGANO joined NEC in 1985, and is currently a manager of the 1st Computers Software Division.

Toshiyuki OHTSU joined NEC in 1986, and is currently a group manager of the 1st Computers Software Division.

* * * * * * * * * * * * * * * * *