Next Generation Grid Storage “HYDRAstor”

SUGIHARA Tomoe, KISAKI Shunsuke, NAKAJIMA Toshiro, MIZUMACHI Hiroaki, KATO Mitsugu

Abstract

HYDRAstor is a grid storage product developed in accordance with the NEC’s three REAL IT PLATFORM concepts of “Flexibility,” “Security” and “Comfort,” in order to resolve serious issues surrounding storage in recent years. This paper introduces the state-of-the-art grid storage architecture of HYDRAstor and introduces its core technologies, Dynamic Topology, DataRedux and Distributed Resilient Data.

Keywords

data, grid storage, backup, restore, duplicate elimination, scalability, virtualization, high reliability, autonomy, disk backup

1. Introduction

The environment surrounding storage is becoming increasingly complex especially in the last several years because of concerns such as the massively increasing amount of digital data, the mission-critical nature and globalization of businesses, mandatory corporate governance and security. The market has been demanding a storage product that can satisfy the complex and diverse range of user requirements, such as superior processing capacities, better cost performance, convenient operational management, dynamic scalability and disaster recovery. At NEC, we developed the next generation grid storage named “HYDRAstor” which realizes our REAL IT PLATFORM concept in order to respond to such market needs. This paper provides an overview of the HYDRAstor and introduces its core technologies.

2. Overview

The HYDRAstor is next generation grid storage developed with technologies invented by NEC Laboratories America, Inc. The HYDRAstor has the following features:

(1) Grid Architecture

This enables dynamic expansion of performance and capacity. Since it is a totally autonomous distributed system, it operates at the most optimum configuration at all times, offering simple configuration management with low operational management costs.

(2) Single Storage Pool

Storage areas distributed and allocated across multiple locations are virtualized and integrated as a single storage pool.

(3) Policy-Based Autonomous Data Operations

Policies such as data storing methods, data protection level and retention periods can be set. Data are operated autonomously based on set policies.

HYDRAstor supports NFS/CIFS, the de facto standard protocol for network file sharing, making it possible to use the system as a filer from UNIX, Linux, Windows and others from a wide range of platforms. It is also possible to configure settings or monitor the status of HYDRAstor from a web browser, without installing any proprietary software. A simple example of HYDRAstor system is depicted in **Fig. 1**.

The core technologies that comprise the foundation of these features are introduced in the next section.

3. The Core Technologies

3.1 Dynamic Topology

HYDRAstor is comprised of two types of nodes, namely, ac-

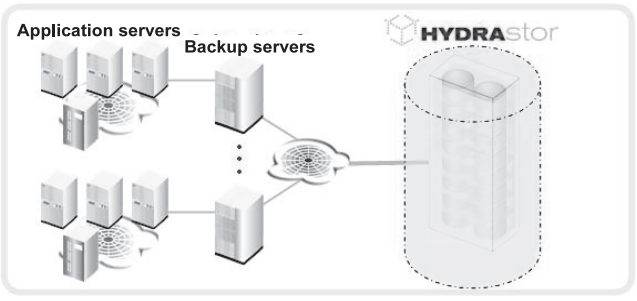

Fig. 1 A simple example of HYDRAstor system.

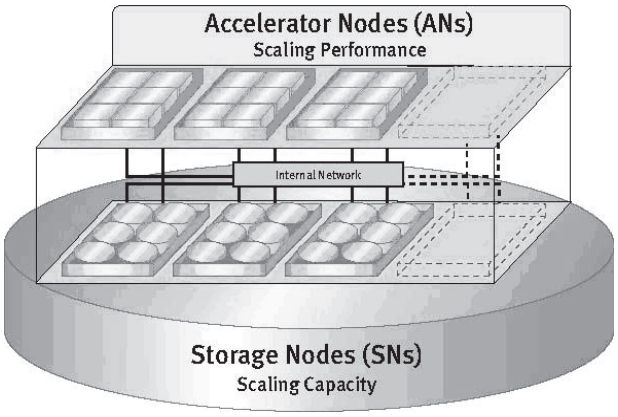

Fig. 2 Grid storage architecture.

celerator node, which processes data requests and storage node, which actually stores data blocks (**Fig. 2**). The performance and capacity of a HYDRAstor system can be expanded dynamically by adding more accelerator nodes and storage nodes respectively (**Fig. 3**). Nodes can be added to the system at any time and without suspending system operations, regardless of where the data storage location may be. Furthermore, added nodes are automatically recognized internally by the system and the distribution of data is re-allocated autonomously to the optimum configuration so that bottlenecks do not arise.

Thanks to this Dynamic Topology technology, the following operational management tasks:

•Expansion of capacity to accommodate increased amount of data.

•Expansion of performance to accommodate increased amount of data traffic.

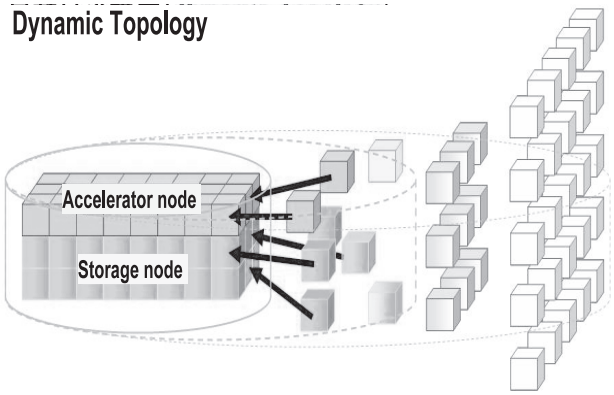

Fig. 3 Dynamic Topology.

•Identifying and removing performance bottlenecks.

•Replacement of nodes upon failures,

which had in the past been extremely complex, can be performed in a simple and easy manner, potentially resulting in significant reduction of management costs.

3.2 DataRedux

DataRedux, a duplicate elimination technology unique to HYDRAstor, checks for duplicates of loaded data. It contributes to high throughput and high cost performance by significantly increasing the data storage efficiency by preventing from loading data that is already loaded on a storage node.

DataRedux intellectually divides data into variable length chunks in order to maximize the detection of duplicates. This enables maximum detection of duplicate data, which cannot be detected by fixed length data division methods (**Fig. 4**).

According to the results of the tests conducted so far, the data reduction ratio (effective stored data amount / physically stored data amount) for backup purposes has been verified to be between 20 and 50. Since the nature of the backup operation causes repeated saving of duplicate data, the data reduction ratio increases each time the data is backed up.

This duplicate elimination technology significantly reduces the traffic to the physical disks as well as the amount of disk space making it possible to perform disk backup at high speed and low cost. Furthermore, by applying this technology to the remote data replication, the transferred data can be further compressed compared to the previous situation where all updated portions had to be sent without data reduction. The amount of transferred data to remote sites can then be significantly reduced, making it possible to perform remote replication with at low speed lines or narrow bandwidths.

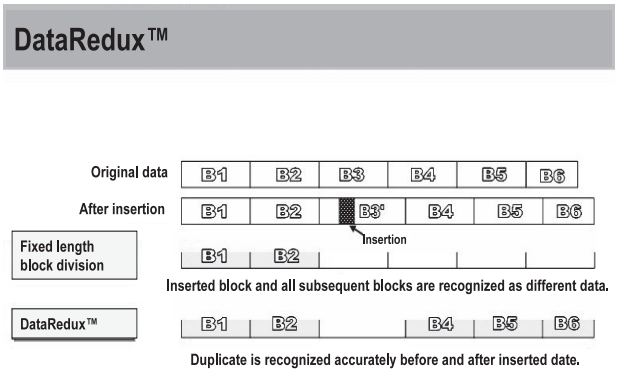

Fig. 4 DataRedux.

3.3 Distributed Resilient Data

When the duplicate elimination technology described in the previous section is used, a single data block can be shared by multiple data. If a data block is lost in such cases, then all data referencing such a data block are impacted and this impact can potentially affect a wide range of data. In order to realize even more robust reliability than conventional RAID (Redundant Array of Independent Disks) systems, Distributed Resilient Data are employed in HYDRAstor; Saved data blocks are further divided into fragments, parity fragments are added before being distributed and stored across multiple storage nodes in order to increase reliability.

Fig. 5 depicts an example of an original data block divided into nine fragments where three parity fragments have been added. In this example, twelve fragments are scattered and distributed across four storage nodes. If a maximum of three out of the twelve of these fragments are lost simultaneously, the original data can still be restored. The reliability of our Distributed Resillient Data is superior to RAID6, which can withstand simultaneous failures of two hard disk drives (HDDs). Furthermore, the redundancy can be set by users, for example according to the importance of stored data, allowing administrators to build and manage the system in a flexible manner.

When some failures occur in HYDRAstor, the failed part is detected automatically and a redistribution process starts in the background. For this reason, there is no cumbersome management work required for administrators. This redistribution procedure is even carried out across multiple storage nodes that have adequate processing capacities without creating any overhead that can encumber existing processes.

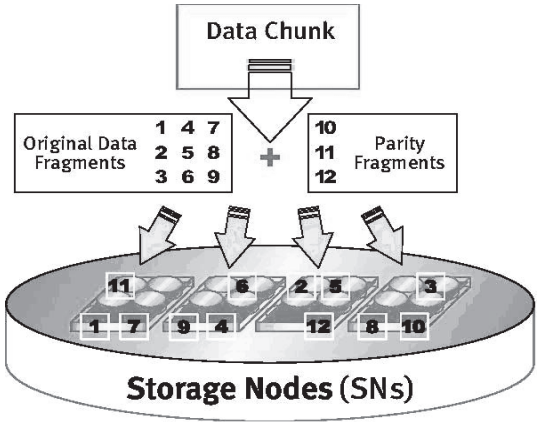

Fig. 5 Distributed Resilient Data.

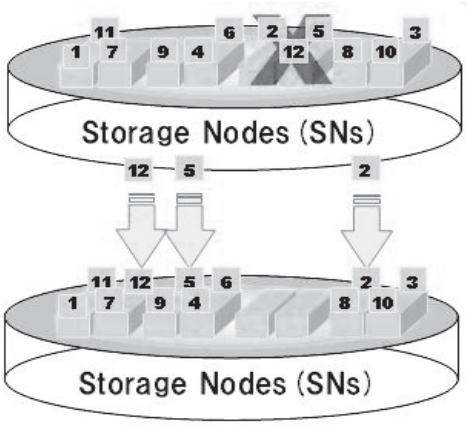

Fig. 6 Autonomous data recovery.

Fig. 6 shows how fragments 2, 5 and 12, which had been lost through a failure, have been detected immediately, then reconfigured and redistributed to other storage nodes.

In conclusion, our distributed resilient data technology has great advantages in reliability and availability compared to conventional disk storage products and reduced TCO (Total Cost of Ownership) upon unfortunate failures.

4. Use Case

By using our core technologies described in the previous sections, HYDRAstor offers not only many benefits such as a high cost performance through duplicate elimination and drastically high reliability by Distributed Resilient Data, but also offers extremely simple implementation and operation at the same time. **Fig. 7** depicts an example of a system used to provide integrated backup for the enterprise data of SAN and corporate organizational data of NAS. In this example, it does disaster recovery by replicating a remote backup site on the HYDRAstor via IP network. HYDRAstor enables such an integrated backup system very easily just by deploying them on both sides of the IP network. Once installed, nodes can be added or disks can be replaced easily as required. Unlike conventional systems, it is no need to perform tedious work, such as adding a tape library to extend the backup processing capacity as the data amount increases or transporting tapes to store them at remote locations; all such issues are resolved by HYDRAstor.

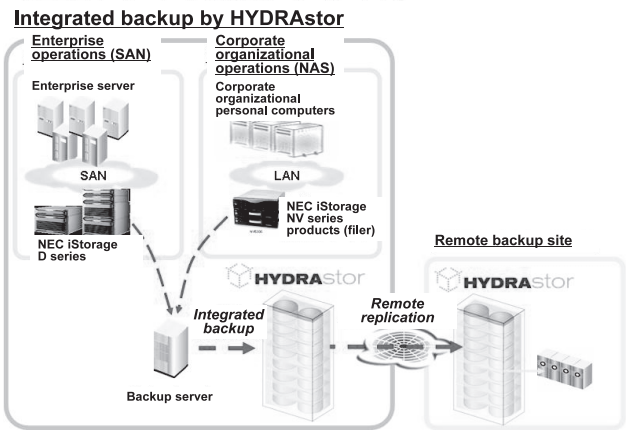

Fig. 7 Use Case

5. Conclusion

In this paper, we have provided an overview of the next generation grid storage, "HYDRAstor" and introduced underlying core technologies. We intend to keep improving and extending features to add capability such as support for primary storage use, data distribution across WAN, and so forth. We are delighted to meet the needs of our customers in a flexible manner.

* UNIX is a registered trademark of the Open Group in the United States and other countries.
* Linux is a registered trademark or trademark of Dr. Linus Torvalds in the United States and other countries.
* Windows is a registered trademark of Microsoft Corporation in the United States and other countries.

Authors' Profiles

SUGIHARA Tomoe
1st Computers Software Division,
Computers Software Operations Unit,
NEC Corporation

KISAKI Shunsuke
Engineering Manager,
1st Computers Software Division,
Computers Software Operations Unit,
NEC Corporation

NAKAJIMA Toshiro
Product Manager,
1st Computers Software Division,
Computers Software Operations Unit,
NEC Corporation

MIZUMACHI Hiroaki
Vice President,
NEC Corporation of America

KATO Mitsugu
Chief Manager,
1st Computers Software Division,
Computers Software Operations Unit,
NEC Corporation

●The details about this paper can be seen at the following.
Related URL:
http://www.necam.com/storage/HYDRAstorWorks.cfm