ETERNUS DX Series of Disk Storage Systems

● Tsuyoshi Kumano

In recent years, information and communications technology (ICT) systems have become more complex and have been handling increasing quantities of data, resulting in increased workloads for ICT system administrators. The provision of efficient administrative functions is thus becoming an important design consideration for storage devices, after the basic function of storing data reliably. Owing to the growing interest in environmental concerns, people have recently started to choose ICT products not only for the functions they offer but also depending on whether or not they include environmentally friendly features. This paper describes how these needs are addressed by the efficient administration features incorporated into the ETERNUS DX series of disk storage systems and the environmentally friendly way in which they are produced.

1. Introduction

In recent years, storage devices have had to store much larger quantities of data, and their operation and administration costs have increased owing to the growing complexity of information and communications technology (ICT) systems. When a storage system is being designed, it is important to consider not only its basic functions of storing data reliably and allowing it to be accessed at high speed, but also the extent to which it resolves operational issues.

Against this background, Fujitsu has brought out the ETERNUS DX series of disk storage systems, which are based on consistent design concepts across the entire range from inexpensive low-end models to high-end models for mission-critical tasks. The series features the RAID (redundant array of independent disks) technology and supports dynamic reconfiguration functions that reduce the operating overheads by allowing the device configuration to be changed without taking it offline. It also uses a technique called "thin provisioning" to facilitate substantial reductions in the workload associated with the up-front design of the system configuration and reduce equipment costs by making it possible to add disk drives only when they actually become necessary.

This paper discusses the dynamic reconfiguration functions: RAID migration, logical unit number (LUN) expansion, and logical device expansion (LDE). It then introduces the thin provisioning technique. Finally, it describes how the system's environmental impact is reduced by using energy-saving features and a battery-free design.

2. Features

2.1 Overview

The ETERNUS DX series achieves a high level of data integrity by combining a redundant component architecture and RAID technology, which uses multiple disk drives to compensate

for disk faults. It features a thoroughly fault-tolerant design that can continue operating without loss of data when various hardware faults occur, including disk faults. It includes a line-up of models ranging from low-end models to enterprise models with a hundred times as much capacity and approximately ten times the performance, as measured using the SPC-1 benchmark **(Figure 1**).

2.2 Reduced operating workload by dynamic reconfiguration functions

In a disk array device, a configuration definition called provisioning is needed to provide servers with services. Provisioning basically refers to the preparations that are made in order to provide a service. It roughly comprises two steps: RAID group creation and volume creation.

1) RAID group creation

A number of disk drives are selected from among those installed in the equipment and they are used to define a set of disk drives called a RAID group. Disk drives that belong to a single RAID group collectively store data in such a way that if any of these disk drives fails, it is still possible to recreate the data on the faulty disk from the data stored on the other disks.

2) Volume creation

A region comprising part (or all) of a preconfigured RAID group is reserved. Here, a volume is defined as storage that can be accessed from a server and is recognized as a "disk" by the server's operating system (OS). The entire storage device comprises a set of multiple disks.

The RAID group configuration and the relationship between volumes and RAID groups are generally fixed on general disk storage systems. Thus, to change a configuration that has already been defined, the system must first delete the RAID groups and volumes that have already been configured and then make a new configuration. This usually inevitably involves prolonged interruption of business while this work is in progress, which leads to increased operational overheads.

The ETERNUS DX series has three functions to make configuration changes less burdensome. All operations are performed without the awareness of the server OS, so changes can be made dynamically while work continues as normal. RAID migration moves volumes between RAID groups, LUN expansion expands the capacity of a volume by connecting it with free space in the same RAID group or

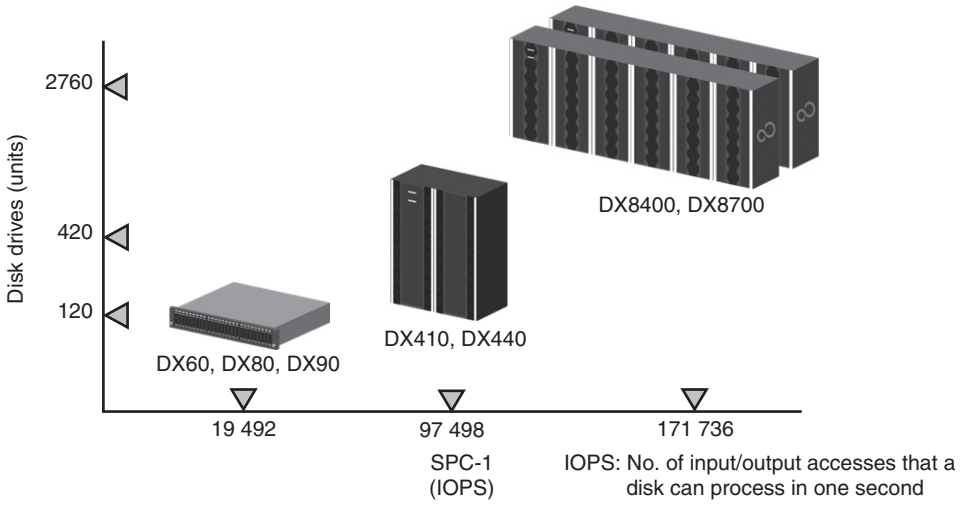

Figure 1
Performance comparison.

in another RAID group, and LDE expands the capacity of a RAID group by adding a disk drive to the RAID group. If these functions are used in combination, it is possible at a later date to adapt to situations where it becomes necessary to modify the storage device configuration owing to changes in business activities or the like.

2.3 Thin provisioning

Thin provisioning is a mechanism that was introduced to greatly improve the operation and management of storage devices. The term was coined to draw a contrast with conventional provisioning, which is now called "fat provisioning" (**Figure 2**), where all the regions of a volume are pre-allocated from contiguous regions of the RAID group and remain fixed thereafter.

Fat provisioning requires ICT administrators to perform the difficult task of deciding the required volume sizes in advance by anticipating the system's future structure, so they tend to be overcautious and create larger volumes than actually needed. This results in the provision of more disk drives than necessary and tends to increase the final cost of introducing the equipment.

In thin provisioning (**Figure 3**), instead of disk space being allocated when the volume is created, it is allocated dynamically when the data is actually written. This eliminates the need to allocate the disk space in advance. It is achieved by interspersing a layer called a storage pool between the volumes and RAID groups. A storage pool consists of one or more RAID groups, and the space in these RAID groups is all lumped together to form the storage pool. When data needs to be written from the server to a volume, it is split into segments of a suitable fixed length and dynamically allocated to the volume from the storage pool in segment units. Segments are allocated from the storage pool so as to distribute them across each RAID group. Distributing the segments in this way means that the disk drives can be expected to operate in parallel, which contributes to increased performance.

From the perspective of the server OS, the volume size is fixed to the size defined when the volume was created, whereas the space actually used by the disk drive consists only of parts to which data has already been written. The term thin provisioning reflects the small actual volume sizes compared with their defined sizes (**Figure 4**).

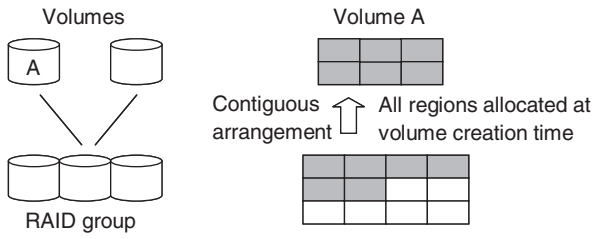

Figure 2
Fat provisioning.

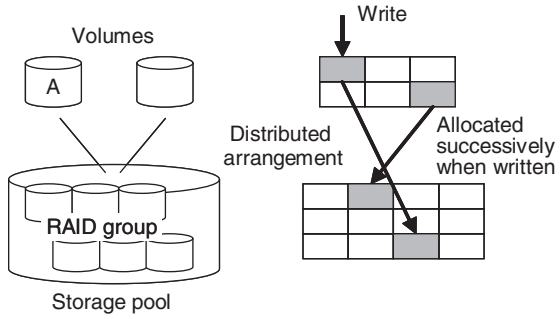

Figure 3
Thin provisioning.

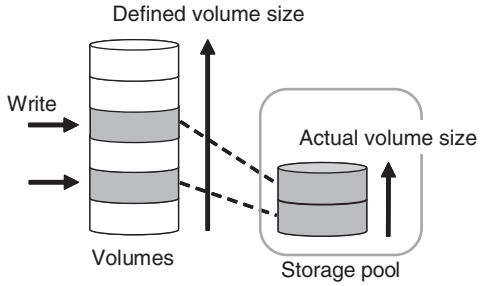

Figure 4
Volume mapping.

Thin provisioning substantially reduces the burden on ICT administrators. When there is insufficient space left in the storage pool, additional disks can be purchased and added to the storage pool without any interruption to business. This approach eliminates the costs associated with having to predefine volumes with an adequate safety margin for estimated future requirements (**Figure 5**), which allows resources to be invested efficiently while the actual status of business is monitored.

When a new RAID group is added to the storage pool, variations in access performance may occur owing to the uneven distribution of data between the new and old RAID groups. The ETERNUS DX series has a function for dynamically equalizing data among RAID groups to eliminate this unevenness.

Furthermore, since the thin provisioning manages the RAID group regions in segment units, ICT administrators are released from the burden of having to manage these regions. In the conventional method, the repeated deletion and creation of volumes in variable lengths results in the free space in the RAID group being fragmented, which can sometimes make it impossible to create a volume even though the total available free space is sufficient. Since storage pool regions are allocated and released in fixed lengths, free region splitting does not occur in principle.

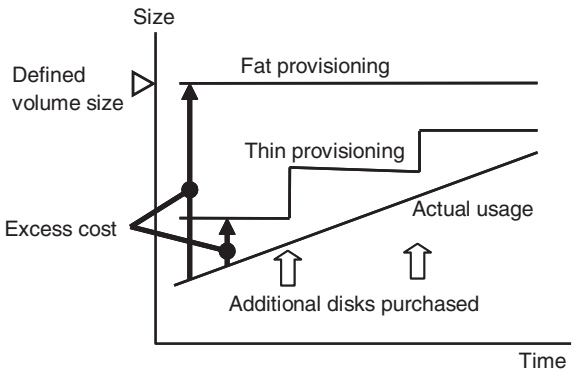

Figure 5
Excess costs of fat provisioning.

Thin provisioning is thus a very effective means of alleviating the burden on ICT administrators and reducing equipment costs. As described below, it is also expected to be an effective means of reducing power consumption. Thin provisioning is available in the ETERNUS DX400 and DX8000 models.

2.4 Reduced power consumption

The ETERNUS DX series has an eco-mode based on massive array of idle disks (MAID) technology, which switches the disk drive motors on and off while the array is operating. The motors are switched on at pre-scheduled times, or when data is accessed, and are switched off if there have been no accesses during a fixed time period, thereby reducing the electrical power consumed by the equipment.

Disk drive motors are normally run continuously while the device is switched on. The storage device cannot predict which disk drive will be accessed next, and it can take a few minutes for the motor speed to stabilize. Consequently, the motors must be kept running constantly to guarantee a fixed response time to access from the server. However, there have recently been a growing number of cases where data backups are stored on disks instead of on tape in conventional tape-based archives. Since backups are often performed at predetermined times such as at night, the eco-mode scheduling can be easily applied.

The ETERNUS DX60, DX80, and DX90 have been designed for reduced power consumption. A more efficient power supply module and an optimal control scheme for the cooling fan's rotation speed have led to an 8% reduction in the power consumption per device compared with conventional equipment.

The thin provisioning mechanism also contributes to reduced power consumption. When a volume is created through conventional fat provisioning, a disk drive of equal size must be

prepared in advance. Since the amount of space on a volume that is actually used to store data generally increases over time, it is necessary to supply power continuously to disk drive regions that are initially unused.

In thin provisioning, additional disk drives can be installed as they become necessary to handle the growing amount of data to be written, so it is not necessary for all the disk drives expected to be needed in the future to be operated right from the very beginning. This leads to lower total power consumption over the equipment's entire operating life.

2.5 Reduced environmental impact

A disk storage system has cache memory to temporarily store the data to be written in order to achieve higher speeds. A write command from the host is completed as soon as the data has been stored in the cache, instead of waiting until it has been written to the disk drive. Therefore, to prevent write data from being lost in the event of a power outage, a mechanism for protecting the data in the cache is needed.

An ordinary disk storage system does this by saving data in a cache powered by an onboard battery. However, this battery unit has a limited lifetime and requires periodic replacement (once every three years), which has a large impact on the environment. For reduced environmental impact, the DX60, DX80, and DX90 feature a new method where the cache memory contents are saved in nonvolatile memory (NAND flash) using electrical power from an electric double-layer capacitor (SCU: System Capacitor Unit) instead of a conventional battery. The SCU has a long lifetime, so it does not require replacement during the entire lifetime of the storage device.

3. Conclusion

The ETERNUS DX series of disk array devices has energy-saving features and dynamic reconfiguration functions to support ICT administration and meet growing demands to not only store data reliably, but also reduce operation and maintenance costs and environmental impact. Fujitsu will continue to develop the ETERNUS DX series taking into account new demands such as these.

Tsuyoshi Kumano
Fujitsu Ltd.
Mr. Kumano is engaged in the planning and development of disk array devices.