

An Analysis of the Storage Virtualization and the New Mode of Capacity Expansion in Colleges and Universities

Yong Yan

Modern Educational Technology Center
China Jiliang University
Hangzhou, China
yy@cjlu.edu.cn

Hongtao Cai

Modern Educational Technology Center
China Jiliang University
Hangzhou, China
cht@cjlu.edu.cn

Abstract—It has become a trend for the construction of storage systems to construct a unified storage platform architecture based on a variety of storage virtualization technologies. The present author analyzes and researches the technical characteristic of storage virtualization and the mechanism of thin provisioning. On this basis, the present author offers a solution to the problem of storage system construction in colleges and universities according to the characteristics of the campus data center construction.

Keywords—*unified storage; storage virtualization; capacity expansion; thin provisioning*

I. INTRODUCTION

In recent years, with the deepening development of the informationization and digital campus construction in colleges and universities, the types and quantities of the campus intranet resources are expanding rapidly. In order to provide better service for teachers, students and relevant departments, and to bring the dispersed network resources under centralized storage and unified management, the colleges and universities has begun to construct a unified campus data center step by step.

Storage system is the carrier of the data, so the construction of the storage system is undoubtedly an important foundation work in the process of the campus data center construction. Because of the accumulation of previous construction, the original storage devices were characterized by dispersion and heterogeneity, which has led to low efficiency and high costs in managing the storage system. At the same time, low storage resources utilization has led to waste in the investment. The introduction of storage virtualization technologies to the storage system construction in colleges and universities can be a quite good solution to these problems. The construction of a unified storage platform architecture based on a variety of storage virtualization technologies masks the heterogeneities of the original storage devices and integrates the original dispersed storage resources easily, which improves the storage resources utilization and provides a unified management interface for administrators.

On the other hand, with the sustained and rapid growth of a variety of teaching resources and public resources on the campus intranet, the campus data center is facing the pressure from rapid expansion of data. Mass data would

inevitably lead to a shortage of storage resources. The traditional storage systems are characterized by poor expansibilities. The capacity expansion is risky and involves complex operations. The construction of a unified storage platform architecture based on a variety of storage virtualization technologies is characterized by high expansibility, and it is pretty easy to expand the storage capacity online continuously. In addition, in recent years, the rapid development of the thin provisioning technology has made the capacity provision a more rational and efficient task, and online capacity expansion has become more simple, convenient and safe.

Based on the analysis of the storage virtualization technologies, the present author proposes an optimized solution to the construction of campus data center storage system — the application of storage virtualization technologies to the construction of a unified storage platform — in accordance with the characteristics and types of the data resources in colleges and universities.

II. VIRTUALIZED STORAGE SYSTEMS

A. Infrastructures of Storage Systems

The storage technologies have existed for a long time. The following are the most commonly used technologies in the storage industry: DAS (Direct Attached Storage), NAS (Network Attached Storage), SAN (Storage Area network).

DAS refers to the storage device directly attached to the application host with a SCSI cable or FC (Fiber Channel) in between, which is the traditional method used in the original storage systems. A leap from the internal storage of the application host to the storage subsystem is realized in DAS, but obviously its disadvantages include poor expansibility, low resources utilization and poor manageability.

NAS, which has a private file system, provides file sharing service to the application hosts connected to the network with some protocols (such as NFS, CIFS). NAS is connected to the application hosts through the TCP/IP network, so it can be constructed on the existing campus intranet. In addition, more and more NAS storage manufacturers have implemented the MPFS/MPFSi protocols, so almost all data can be transmitted via FC or IP network when making access to the NAS resources except for the control instructions and meta data which are still

transmitted via IP network. The data access becomes more efficient than before in this way. However, getting access to NAS system need to go through format conversion of the file system, so it is not suitable for block level data transmissions, especially those that need raw devices. For example, the application of some database system would be a problem.

SAN is relatively new. It is a network storage solution which connects the application hosts and storage devices through the private network. SAN can be divided into two types according to different protocols and connections: FC SAN connected through FCs using the SCSI protocol, IP SAN connected through Ethernet cables using the iSCSI protocol. SAN provides block level data transmission, which is better than NAS in this respect. Also because the transmission bandwidth of the private network is generally better than that in NAS, the data transmission capacity has been improved. However because there is no file system, data cannot be shared among different businesses just like NAS, and data sharing must be implemented with special software. On the other hand, in the ordinary course of events, users need to purchase expensive special switches to build a private storage switching network, and the cost is high.

B. Storage Virtualization Technologies

Storage virtualization technologies refer to the process of mapping heterogeneous physical storage devices to virtual storage pools which are built in accordance with some virtual storage architecture. The application hosts will get access to the logical volumes built in the virtual storage pools instead of the physical storage devices.

Storage virtualization solved the problem of unifying heterogeneous storage devices. As a result, users do not have to give up the existing storage devices when planning new storage system, and they do not have to be limited to using the same brand of storage products from the beginning to the end. So the architecture of the unified storage platform emerges.

In addition, the unified storage systems based on storage virtualization abstract the underlying physical storage devices and provide virtualized logical storage capacities for the application hosts. Meanwhile, the unified storage model centralizes the dispersed storage resources, so the hosts can have access to the logical volumes without regard to physical storage devices and their locations. On the other hand, storage virtualization and unified storage model make it easier to manage and maintain the storage system, and they are characterized by high expansibility, which is quite convenient.

Virtualization of storage can be divided into three levels according to different implementation methods: host-based virtualization, storage-subsystem-based virtualization and network-based virtualization.

Host-based virtualization is implemented by appending a virtual middleware layer in the operating system. This layer is commonly referred to as LVM (Logical Volume Manager),

which maps the physical storage devices to logical volumes where the file system is built.

Storage-device-based virtualization is implemented by the processor within the storage device, and the function of virtualization is directly integrated into the storage hardware.

Network-based virtualization refers to the paths from the hosts to the storage devices. NAS and SAN are examples of this model.

III. MODES OF STORAGE CAPACITY EXPANSION

In the context of the ever-increasing data expansion of the campus data center in colleges and universities, it is almost impossible to foresee the storage demand accurately. When the existing storage resources are exhausted, it is inevitable to expand the storage capacity.

A. Capacity Expansion Modes Based on Storage Virtualization Technologies

Early traditional storage systems do not support dynamic expansion of volumes, and thus cannot support online expansion. Storage systems using the technologies of storage virtualization support uninterrupted expansion of storage capacity online, which meets the needs of the sustaining and unpredictable storage expansion.

The virtual storage pools built based on storage virtualization technologies not only solve the problem of heterogeneous storage devices, but also provide online dynamic allocation of storage capacity, so it made the online capacity expansion very convenient. Administrators can expand the capacity of a virtual storage pool by adding new physical storage devices and mapping them to it. A new LUN (Logical Unit Number) can be created with the added capacity, thus expanding the capacity of the application. All these operations can be done online.

In the application host, the traditional partition mode of the operating system requires continuous storage space, so the added storage space cannot be directly extended to the existing partition. To expand the existing one, the administrator must rebuild it with special software, and the operation generally cannot be done online.

The previously mentioned host-based storage virtualization technologies such as LVM resolve this problem in the host side. LVM provides an online dynamic allocation of the logical volumes. With the added capacity, a new logical volume can be easily created online. The administrator can also expand the existing logical volume online with the added capacity, thus further expansion of the original partition capacity of the file system based on this virtual layer can be completed.

B. Thin Provisioning Technology

The traditional storage provisioning preallocates adequate physical capacity for each application. How much physical capacity will be purchased is determined by the

current capacity planning. To avoid the storage capacity expansion or reallocation that may lead to business interruption and possible risks of data loss, the administrator often physically preallocates overmuch capacity to applications while planning the storage system. Once the certain storage space is preallocated to an application, it cannot be allocated to another with the traditional provisioning technology. However, these applications cannot make full use of the storage capacity commonly. Therefore, the utilization of the storage resources is very low. This leads to an extreme waste of storage resources and storage costs.

Thin Provisioning technology can solve this problem well. Thin provisioning, a relatively new technology, was first introduced to the storage community in 2002, mainly for the file-system-based virtualization within NAS systems. Then some SAN manufacturers began to provide this function. Thin provisioning technology has been developing with tremendous speed in recent years, and applications based on this technology have also undergone an unprecedented explosive growth.

Thin provisioning is a new technology of storage virtualization, which provides a new mode for the storage allocation and capacity expansion. Fig. 1 illustrates the mechanism of this technology. With thin provisioning, the physical storage space is put into a public storage pool. In the figure, the logic capacity of the public storage pool which is composed of the physical storage space is 100GB. This storage system is connected to three applications: A, B, C, and the applications get 100GB, 200GB, and 300GB capacity respectively virtualized with thin provisioning, which is obviously bigger than the real logical 100GB capacity provided by the physical devices. The applications get a full size virtual space without knowing the actual available space is far from enough. In the figure, the three applications currently occupy 10GB, 20GB, 30GB, and the real remaining logic capacity is 40GB.

Thin provisioning relies on on-demand allocation of data blocks to allocate capacity for applications from the public storage pool. The storage system provides applications with virtual storage capacity, and real physical storage capacity is only dedicated and mapped when data is actually written by the application, which is transparent to the application. The administrator can easily set up the monitoring and alarming system with special software. When the real capacity usage reaches a certain alarm level, new disks or storage devices can be purchased timely, and then added to the public storage pool, thus expanding the physical storage space.

This new technology undoubtedly brings great convenience to dynamic allocation and online expansion of storage capacity. Administrators can make a long term plan (even several years) of larger storage space for the applications, but do not need a one-time purchase of the equivalent amount of physical devices. They can allocate an expected virtual storage space for the application with thin provisioning, and then purchase new physical devices when the physical space is almost full. These operations can be

conducted online and at the same time, they are completely transparent to the application.

Administrator can dynamically allocate virtual storage space of each application according to the actual usage, making the capacity allocation tend to rationalization, and at the same time increasing the utilization of storage resources substantially. The deferred purchase plan brings about the cost of delay, reduces the one-time budget and makes the storage system become green.

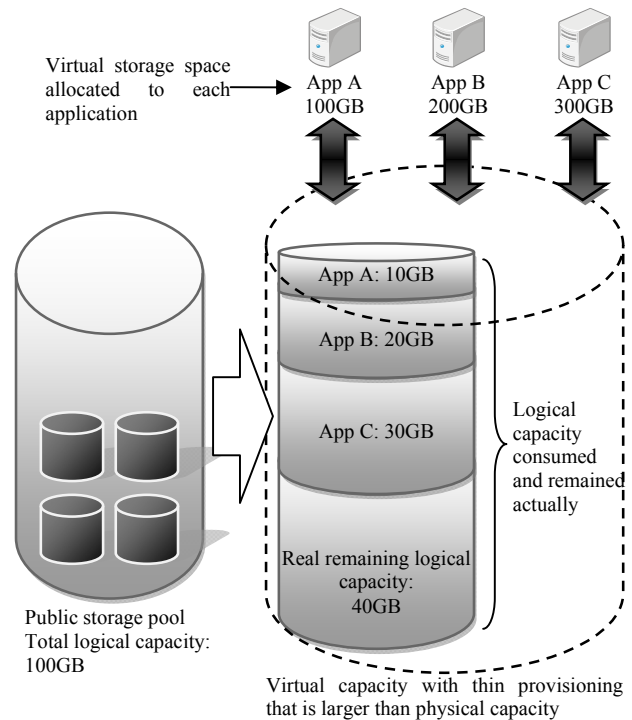


Figure 1. Demonstration of the mechanism of thin provisioning

However, just because thin provisioning provides virtual storage space for applications and allocate real storage space on demand, this technology is not applicable to applications that need to initialize the whole given space.

In addition, the on-demand allocation principle of thin provisioning results in the applications' seizing of the space on first-come-first-served basis, so thin provisioning often has no sensitivity toward the data location of the real physical storage device. But some high performance applications want their file system highly optimized so that the data stream can be directed to the fixed location of physical space. Therefore, such applications cannot get expected high performance in thin provisioning environment. Similarly, some applications that are doing heavy loaded tasks of sequential writing will get reduced performance because thin provisioning leads to the discontinuity of the actual location of data.

Nevertheless, thin provisioning is suitable for enterprise-level applications which will accept small changes in

performance. Therefore, thin provisioning can be widely used in storage systems, as enterprise-level applications, in colleges and universities.

IV. DEMAND FOR STORAGE SYSTEM IN COLLEGES AND UNIVERSITIES AND A SOLUTION

A. Demand Analysis of the Storage System in Colleges and Universities

The construction of the resources platform of campus data center in colleges and universities involves the core businesses that are of prime importance, including campus card, unified identity authentication, central database of digital campus. It also involves the web portals, mail systems, VOD, courseware library of elaborate courses, FTP, file sharing systems, 3D virtual campus, the electronic library and other public information service systems. In addition, it includes the storage resources of campus monitoring systems, network security systems and other business systems.

The features can be broadly summarized as: a wide variety of applications, a great variety of data types, a large amount of data and the high concurrent data access caused by the large quantity of computer clients.

These require stable, secure, efficient and fast mass storage space. Meanwhile, the management mode of multi-machine clusters and centralized servers requires centralized storage. The unpredictability of the data expansion rate also makes heavy demands on the dynamic allocation and online extension of storage capacity.

The application of a unified storage platform can improve the utilization of storage resources and simplify the complexity of system management. The following key points should be considered in the planning: high expansibility, flexible and convenient networking, multi-mode application of SAN / NAS integration, efficient and convenient management and maintenance, the capacity of the storage devices and the price performance ratio.

According to the characteristics and demands above, a unified storage platform architecture based on a variety of advanced storage virtualization technologies will be the best choice in the construction of the storage resources platform of the data center in colleges and universities.

B. Solution to the Storage System in Colleges and Universities

Resources of the campus data center are multifarious, and the referred access methods are diverse, which include block-level data access such as databases, e-mail, and file-level data access such as VOD, FTP, web browsing. Data access operations are also divided into different ranks. There are core businesses of digital campus which need high bandwidth and high concurrency, and internal common public services.

Therefore, the SAN architecture should be the primary consideration, and a unified SAN and NAS storage platform

integration would be a better solution. The application of virtualized storage architectures can further improve efficiency, simplify management, and fully integrate the storage devices. Thus the platform provides data access separately based on file-level or block-level for corresponding different businesses. Fig. 2 shows the integrated storage solution for colleges and universities. What kind and how many physical storage devices should be purchased are based on the size of the college and the demand of the data center, and the stable and high-speed FC-based storage products are preferred. A virtual storage pool is built with virtualization technologies upon the physical storage device groups, and provides data access for the applications through different ways. Databases, e-mail and other block-level data access and core businesses use the SAN-based storage network that provides high bandwidth, high throughput; FTP, VOD, Web and other file-level data access use NAS model that provides fast file-level access for these applications and is compatible with different operating systems.

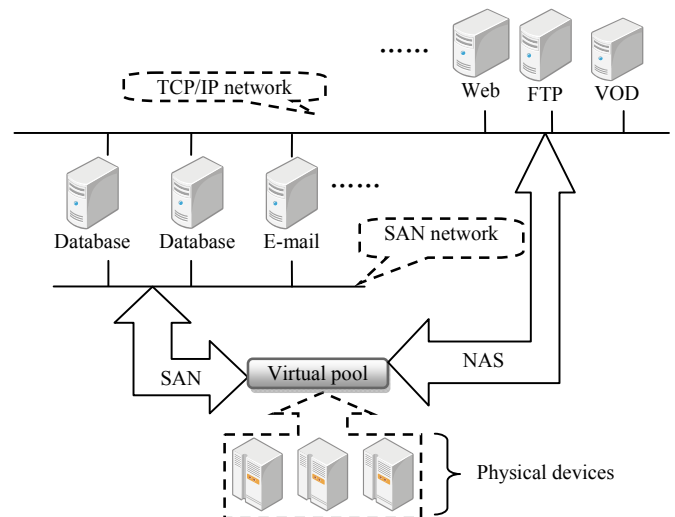


Figure 2. Integrated storage solution for colleges and universities

Because of the unpredictability of the data expansion, administrators should plan the storage capacity respectively for a short-term and a longer period demand in the storage system. They should only purchase the physical storage space which just meet the short-term storage demand, and apply thin provisioning to the storage subsystem, and then allocate reasonable virtual storage capacity for different applications. The operations of optimization and re-adjustment should be taken for the capacity allocation according to the practical situation of the capacity utilization during the actual runtime. At the same time, the application hosts connected to the storage subsystem should use LVM to manage the storage space and to build logical volumes, which provides convenience for the dynamic adjustment operations of the logical volumes in the future.

V. CONCLUSIONS

Virtualized storage technologies and the unified storage platform architecture are the current development trend of the storage industry. With the deepening development of the informationization and digital campus construction in colleges and universities, the campus data center is holding an increasingly high position. So the storage system, as the foundation work of the data center, is increasingly important. With the vision of a global and long-term development, the construction of a centralized storage system with virtualized storage technologies and the unified storage platform architecture has become the first choice in the construction of data center in colleges and universities.

The storage virtualization technologies, thin provisioning, and the storage system solutions for colleges and universities discussed in this paper, as a successful attempt to build a the centralized storage system in colleges and universities, has reached the target mentioned above.

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to the Leading Group of Digital Campus of China Jiliang

University and Dongqi Electronics Co., Ltd. in Hangzhou. They have helped me a lot in the writing of this thesis.

REFERENCES

- [1] Jikui Wang, The road to storage IT professionals: storage from ABC to storage professional. Beijing: Tsinghua University Press, 2009 (In Chinese).
- [2] Shiwen Lu, The technology and application of storage network. Beijing: Tsinghua University Press, 2010 (In Chinese).
- [3] Yu Wang, Xiaofeng Liu and Xingwei Wang, Applications of virtualization technology in the informationization construction of universities. Journal of Frontiers of Computer Science and Technology, 2010,4(4), pp. 353-358, 2010 (In Chinese).
- [4] Shenglong Tan, Research on storage virtualization technologies. Microcomputer Applications, vol. 31, no. 1, pp. 33-38, Jay. 2010 (In Chinese).
- [5] Hongfei Qiu, Research of Thin Provisioning Technologies on Storage. Telecommunications Science, 2010, no. 11, pp. 12-17, 2010 (In Chinese).
- [6] T. Clark, Storage Virtualization: Technologies for Simplifying Data Storage and Management. Addison Wesley Professional, 2005.
- [7] C. Milligan, S. Selkirk, "Online Storage Virtualization: The Key to Managing the Data Explosion," hicc, vol. 8, pp.225b, 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 8, 2002.