Hitachi Virtual File Platform for Content Cloud

Hiroshi Ogasawara
Yuichi Yagawa
Motoaki Satoyama

OVERVIEW: Content data, particularly unstructured data such as images or e-mail, is undergoing tremendous growth. Unstructured data is stored in file format, and while the typical way of dealing with the increase in the quantity of file data in the past has been to connect NAS, the rapid growth in recent years has highlighted issues with the complexity of administration. Hitachi has devised the concept of a "content cloud" as a way of solving this problem and supplies Hitachi Virtual File Platform as a more advanced alternative to conventional NAS. Features include integration of infrastructure within and between sites, automatic consolidation of files at a data center, effective use of hardware and personnel resources, and reduced administrative procedures. Use of this product as a core element in the provision of cloud services also facilitates the adoption of cloud computing.

INTRODUCTION

THE quantity of data handled by corporations tends to grow relentlessly year after year. However, in recent years, data has come to be stored in different ways according to the data types, with the rate of growth in the quantity of file-based and other unstructured data being particularly rapid.

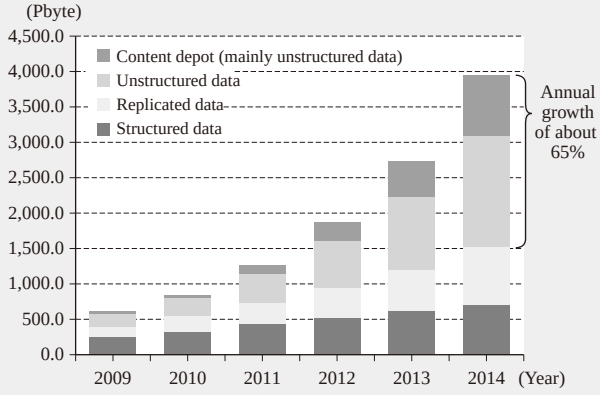
Market research company International Data Corporation Japan classifies data into four groups: (1) structured data, (2) replicated data, (3) unstructured data, and (4) content depots. Here, structured data means data held in databases or used for applications such as ERP (enterprise resource planning), SCM (supply chain management), or CRM (customer relationship management). Replicated data includes replication, backups, and archives. A characteristic of these types of data is that they are mostly generated and stored in data centers.

In contrast, unstructured data and content depots are often located on local storage devices such as NAS (network-attached storage) or file servers. Unstructured data includes text, images, audio, video, e-mail, and CAD (computer-aided design) documents and is typically held in file format. Content depots, meanwhile, store digital content published by service providers such as news, stock prices, financial information, technical information, academic information, comparisons of prices or similar, video, music, and games. Most of this is unstructured data and is produced and stored exclusively in file format.

The quantity of unstructured data is anticipated to grow at an annual rate of about 65% between 2009 and 2014 (see Fig. 1). This will exceed what existing NAS, file server, and other locally installed storage systems will be able to handle adequately, and the situation is getting to the point where the complexity of operating and administering this data can no longer be ignored in terms of its impact on business.

Hitachi has come up with the concept of the content cloud as a way to manage this unstructured data efficiently while reducing the associated system installation and running costs, and has released Hitachi Virtual File Platform*1 (VFP) in Japan, which incorporates file virtualization with the aim of putting the concept into practice.

*1 Hitachi Virtual File Platform is a file storage product sold in Japan market only.



Source: "Analysis of Consumption Model for 2010 Market for Disk Storage Systems in Japan" published by IDC Japan

Fig. 1—Growth of Japanese Disk Usage by Data Type. Average annual growth is expected to increase between 2009 and 2014.

This article gives an overview of the content cloud and describes the features and functions of the VFP, which include support for the cloud that goes beyond that offered by existing NAS products.

CONTENT CLOUD FOR MANAGING RAPID INCREASE IN DATA

Overview

Described briefly, the content cloud is a system for storing and utilizing data in a safe, reliable, and convenient manner, without concern for the type of data and its other characteristics.

In the past, the practice has been to provide a separate file server or NAS unit for each of the various different types of content (files), such as files, e-mail, or images, and to create backups and perform other management tasks for each of these units independently. As a result, operations have become complex. In a content cloud, the data is administered centrally within the cloud environment and tasks such as making backups can be consolidated. This simplifies operating procedures, and resources that were wastefully spread across these different servers can be centralized to make their use more efficient (see Fig. 2).

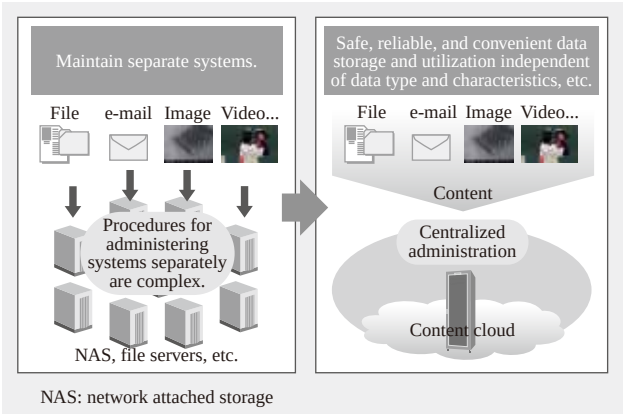


Fig. 2—Overview of Content Cloud.
Whereas past practice was to add NAS units as data quantities increased, which tended to cause systems to become divided, the content cloud consolidates data access from applications.

Step-by-step Implementation of Content Cloud

When moving to a content cloud from the past practice of maintaining separate systems, an alternative to making this shift all at once is to adopt a step-by-step approach. That is, preparing for the cloud involves first consolidating the servers, NAS, and other devices at a site, and then automating administration and shifting to tiered data management. This facilitates the transition to a cloud environment while allowing users to consolidate their in-house systems efficiently.

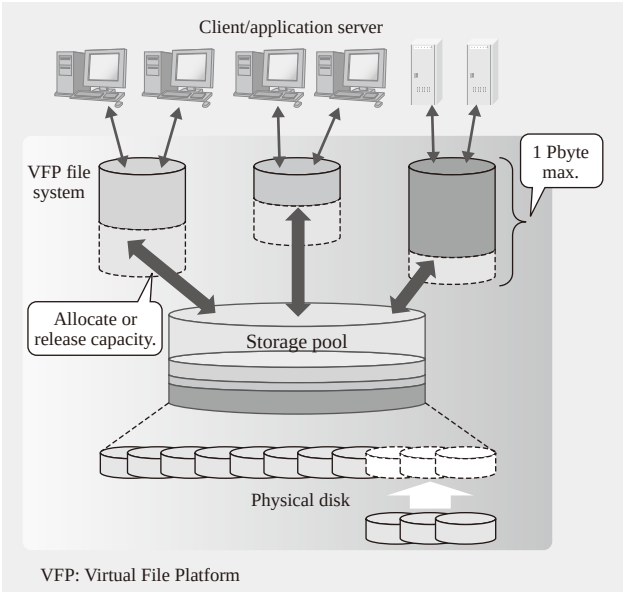


Fig. 3—Efficiency Improvements Due to Capacity Virtualization.
The work associated with preliminary capacity design is reduced by predefining sufficiently large virtual capacities. Also, the physical capacity used by file systems can be allocated automatically.

VIRTUAL FILE PLATFORM FOR CONTENT CLOUD

Through the VFP, Hitachi has made it possible to implement a content cloud incrementally. The following sections describe the relevant functions.

High Level of Expandability Allowing Consolidation

A system intended to consolidate a large number of NAS, file servers, or other devices requires adequate performance and capacity. The latest VFP delivers approximately twice the performance of the previous model, and with a maximum capacity of 15 Pbyte and support for file system sizes up to 1 Pbyte (60 times the previous maximum), it has the capacity needed for NAS and file server integration.

The capacity virtualization function means that it is not necessary for the initial disk capacity to be equivalent to the specified file system size. Through interoperation with Hitachi Dynamic Provisioning function of Hitachi Adaptable Modular Storage 2000 series, it is possible for the storage pool to be shared by a number of file systems. As capacity is allocated from the pool automatically when data is written to a file system, capacity can be provided to satisfy growing data quantities in a timely manner. Also, because the

system administrator is notified when the amount of storage being used approaches the physically installed disk capacity, the purchase and retrofitting of additional disks can be done as and when required to help keep storage costs under control. When a file system or large amounts of file data are deleted, such as when a department is moved, the space freed up by deletion is made available and can be allocated to other file systems (see Fig. 3).

Consolidation of Storage with Different Access Protocols onto One Unit

VFP supports the block storage access protocols Fibre Channel and iSCSI (Internet Small Computer System Interface) as well as the commonly used NAS protocol NFS (network file system), CIFS (common Internet file system), and FTP (File Transfer Protocol). The administration screens have the same design with Hitachi Adaptable Modular Storage 2000 series to allow both products to be administered from the same screen. This allows access from applications using a range of different access protocols to be consolidated onto a single unit to reduce both storage costs and running costs.

Integration with External Software

The VFP works with software such as VMware*2 and Hitachi Command Suite to allow more efficient administration. With VMware, it provides a plug-in that can be used to create data stores on the VFP from the VMware vSphere client GUI (graphical user interface), store VM (virtual machine) image files, and perform backup and restore. Backup scheduling is also supported, and the VFP allows VMware to be used without having to take account of storage.

Online Data Transfer Function to Facilitate System Migration

The VFP also provides a function for the online transfer of data from NAS, files servers, and other devices. Using file virtualization, once the VFP is connected to the LAN (local area network) and the transfer settings completed, access to all data becomes available as soon as the user initiates the transfer. This applies whether or not the data has been transferred. Accordingly, operations, which in the past would have needed to be resumed after the transfer completed can now be resumed immediately after the transfer begins, significantly reducing the scheduled down time.

*2 VMware and VMware vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.

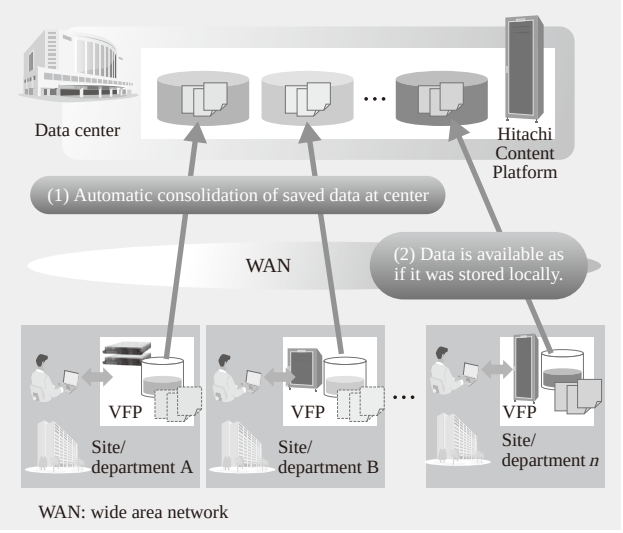


Fig. 4—Automatic Data Consolidation Using Cloud on-Ramp. Backups and archiving to the center are performed automatically. File virtualization of site or departmental data makes its physical location transparent to users.

Cloud on-Ramp Solution

The VFP's file virtualization function allows data written to the local VFP to be backed up or archived automatically on Hitachi Content Platform at the data center (see Fig. 4).

The local VFPs interoperate with the data center to provide an on-ramp to the cloud, which transfers data from the VFPs to Hitachi Content Platform in accordance with predefined policies specified by the administrator. If a file or other data on the VFP is not accessed for a specified period of time, it is converted to a stub based on the policy settings and the local copy of the file is deleted. This ensures that disk capacity on the VFP is used efficiently. It also lets systems operate without concern for local disk capacity because, while important files that are accessed frequently remain on the local VFP where they can be accessed with high speed, data that needs to be kept even if it is not accessed is stored on Hitachi Content Platform at the data center. If a file stored on Hitachi Content Platform that has been converted to a stub on the local VFP is accessed, the data is restored to the VFP so that it is subsequently available for local high-speed access. If the restored data becomes inactive once more, it is converted back to a stub in accordance with the policy, just like any other data, thereby maintaining efficient use of the VFP capacity.

If a VFP is put out of action by a disaster or other local event, the cloud on-ramp solution allows operation to resume as soon as the VFP is replaced. This is because, when a VFP is replaced, it restores its management information from the data center Hitachi

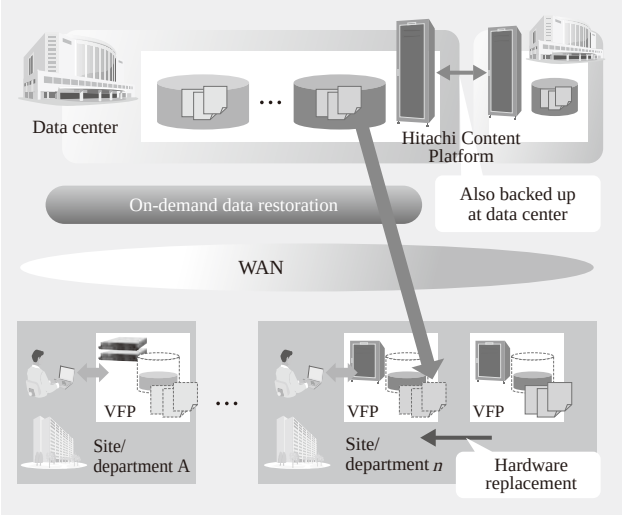


Fig. 5—Quick Recovery of Data Access after Local Disaster. If a local disaster occurs, the cloud on-ramp solution can restore data access as soon as the hardware is replaced.

Content Platform and then progressively restores data to the VFP on an on-demand basis as it is requested. That is, users can have immediate access to the data they require without needing to wait until all of the data has been restored from Hitachi Content Platform (see Fig. 5).

Also, because Hitachi Content Platform at one data center can replicate the data from Hitachi Content Platform at another, the system can cope with disasters at the data center as well as at local sites.

The cloud on-ramp solution can use a multi-tenanting function whereby each local site is allocated a separate area at the data center (see Fig. 6). Using the figure as an example, inter-site security can be maintained by prohibiting site B from accessing backup data from site C. In addition to companies that want to enforce inter-departmental security, this function is also convenient for providers who intend to use the cloud on-ramp solution to supply cloud services to customers. The efficiency of capacity utilization for data at the data center can thus be improved without compromising security by performing tasks such as compression or eliminating duplications as policy-based background operations. Also, because tamper-prevention attributes specified for local files are carried over when the file is stored at the data canter, it is possible to combine safe and secure data storage for each application with more efficient capacity utilization.

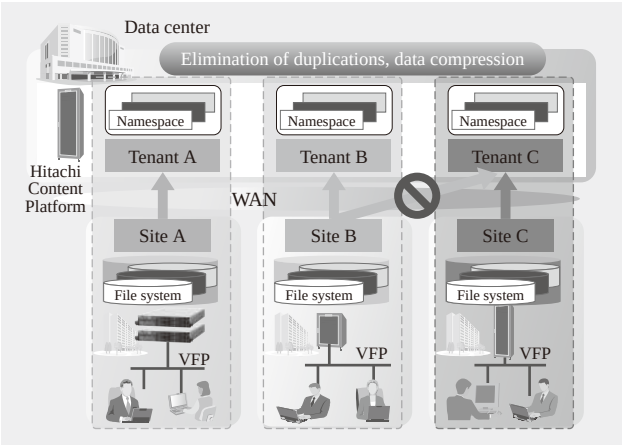


Fig. 6—Secure Data Storage Using Multi-tenanting. The VFPs at each site are allocated tenancies at the data center Hitachi Content Platform to ensure independent security.

CONCLUSIONS

This article has given an overview of the content cloud and described the features and functions of the VFP, which include support for the cloud that goes beyond that offered by existing NAS products.

VFP can use file virtualization to manage the rapidly growing quantities of content data in a content cloud. This makes effective use of hardware and human resources and reduces the administration workload. It is also possible to lower the barriers to entry for users adopting cloud computing by providing cloud services in which this product plays a core role.

Hitachi intends to continue contributing to reducing the cost and improving the efficiency of data use by customers by providing new functions and enhancing the performance of the content cloud.

ABOUT THE AUTHORS

Hiroshi Ogasawara
Joined Hitachi, Ltd. in 1995, and now works at the Disk Array Systems Division, Information & Telecommunication Systems Company. He is currently engaged in the planning and marketing of file storage products. Mr. Ogasawara is a member of the Information Processing Society of Japan (IPSJ).

Yuichi Yagawa
Joined Hitachi, Ltd. in 1991, and now works at the Strategic Planning Office, Yokohama Research Laboratory. He is currently engaged in the R&D strategic planning.

Motoaki Satoyama
Joined Hitachi, Ltd. in 1987, and now works at the Disk Array Systems Division, Information & Telecommunication Systems Company. He is currently engaged in the development and design of file storage products.