

Windows Server 2008 R2 Hyper-V Server Virtualization

● Yutaka Haga ● Kazuhide Imaeda ● Masayuki Jibu

Server virtualization has attracted attention as a means of reducing the costs of system introduction, operation, and management through server consolidation and as a platform for public cloud computing. The Microsoft Windows Server 2008 R2 operating system provides the Hyper-V server virtualization function as a standard feature. Hyper-V uses a microkernel-based hypervisor as an architecture for a flexible, stable virtualization platform. Fujitsu has accumulated technological knowhow and troubleshooting expertise from its experience in combining this Windows Server operating system with the PRIMERGY PC server and testing beta version of the Hyper-V. The knowhow acquired from this experience is available on a website put up by Fujitsu. This paper covers the Windows Server 2008 R2 Hyper-V architecture, virtual disks, virtual networks, file-sharing systems, and high availability functions. It also introduces the architecture and management functions of Microsoft's System Center Virtual Machine Manager, which improves the manageability of the Hyper-V environment.

1. Introduction

Server virtualization has been attracting attention as a means of reducing the costs of introducing and operating business information and communications technology (ICT) systems. Server virtualization is expected to reduce energy consumption, reduce installation space, and enable efficient utilization of server resources. Against this background, a number of vendors are entering the virtualization market with products that incorporate proprietary server virtualization techniques.

Fujitsu has accumulated technological knowhow and troubleshooting expertise from its experience in combining the Windows Server 2008 R2 operating system (OS) with the PRIMERGY PC server and in testing beta version of the Hyper-V server virtualization function provided as a standard feature in the Windows Server OS. The knowhow acquired from this experience is available on a website put

up by Fujitsu.^{1,2)}

In this paper, we introduce the architecture and distinguishing features of Hyper-V and its management functions. We also describe an advanced case study from Mitsui & Co., Ltd., where these functions were used in an attempt to implement a private cloud.

2. Hyper-V architecture

Hyper-V uses a hypervisor type of virtualization architecture,³⁾ which is provided as an extremely small component called Windows Hypervisor. A virtualization platform based on Hyper-V can be configured in a simple manner by adding the Hyper-V role after Windows Server 2008 R2 OS has been installed in the usual manner.

Instead of using a monolithic hypervisor, it uses a microkernel-based hypervisor in that the device driver (Windows driver stack) is embedded in the management OS (parent partition) rather

than in the hypervisor itself (**Figure 1**). This means that drivers that are compatible with the management OS can be embedded easily in the virtualization platform and that the drivers are prevented from affecting hypervisor quality.

We will now touch upon the I/O processing of the guest OS (child partition). From the guest OS, I/O processing such as network and storage operations appear to be direct operations on a physical device, but the actual I/O processing is done by communications between the guest OS and the management OS to pass requests to the device driver of the management OS, where the actual processing is done. In addition, a communication path called a virtual machine bus (VMBus) is used for communicating between the guest OS and the management OS. An I/O request from a virtualization service client (VSC) on the guest OS is sent to the virtualization service provider (VSP) of the management OS through the VMBus, and the VSP works with the device driver of the management OS to execute the I/O process.⁴⁾

3. Virtual disk

With Hyper-V, a file embodied in the virtual hard disk format (a VHD file) can be used as a disk on a virtual machine (VM). There are three

virtual disk formats: fixed size VHD, dynamically expanding VHD, and differencing VHD. Among these formats, the characteristic one is dynamically expanding VHD, where data blocks expand to the size necessary for storage during a data write.⁵⁾ This makes it possible to minimize the initial size of the virtual disk. In addition, unused data blocks can be deleted by optimizing the virtual disk from the management OS, even for areas that have been temporarily expanded since the management OS can identify unused data blocks in the NTFS file system within the virtual disk. By providing these functions, Hyper-V supports a high level of physical disk utilization.

It is also possible to utilize a pass-through connection in the configuration to assign physical disks to the VM. This provides an I/O capability that is substantially equivalent to that of a physical disk and enables the provision of 2 TB or more of disk space.

4. Virtual network

A component that plays an important role in network communications for a VM is a virtual switch. A VM communicates through a virtual switch connected to a virtual network interface card (NIC). When the VM communicates with

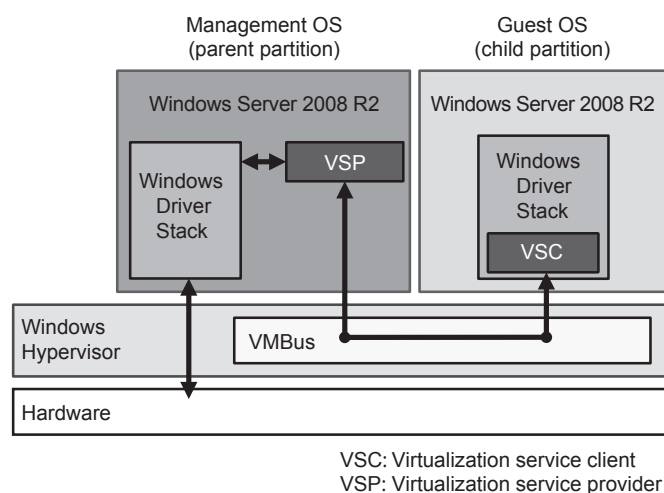


Figure 1
Architecture of Hyper-V server virtualization function.

an external server, it communicates through the virtual switch and a physical NIC in sequence (**Figure 2**). The virtual switch provides standard L2 switch functions, such as switching functions (MAC address learning), virtual local area network (VLAN) functions, and data buffering. The virtual switch is mounted on the management OS as a protocol driver (Microsoft Virtual Network Switch Protocol).

We will now describe the hardware assist function, which lightens the load on the processor by offloading some of the functions performed by the virtual switch to the physical NIC side. Hyper-V uses the virtual machine queue (VMQ) for the offloading.

When a packet is received from an external network in an environment in which VMQ is not used, the virtual switch performs packet routing for each VM on the basis of received packet stored in the buffer of the physical NIC. The MAC address and VLAN ID of the virtual NIC are used in the routing, and the received packet in the buffer is copied and stored in the buffer of each VM prepared by the virtual switch.

When a packet is received from an external network in an environment in which VMQ is used, a queue and buffer are established for each VM on the physical NIC, and the physical NIC

stores the packet directly in the buffer for each VM.⁶⁾ The MAC address and VLAN ID are used to determine the routing by identifying which queue and which buffer to use. Server processor utilization is thereby reduced, and overall system throughput is improved by having the physical NIC itself execute these operations.

5. Hyper-V's high availability and serviceability

Hyper-V uses the Microsoft failover cluster (MSFC) function, a standard Windows Server 2008 R2 OS function, to provide high availability and serviceability.

If an abnormal stop occurs due to a hardware fault in the physical server or to an abnormal termination or hang of the management OS (Windows OS), the VM can continue operating through "failover" of the VM to another physical server. Hyper-V also provides a live migration function that enables a transition to another physical server without halting and restarting the VM, enabling dynamic physical server maintenance and dynamic VM reallocation.

We now explain how live migration works. While the VM is running, the live migration function creates another VM on the server at the transition destination and duplicates the

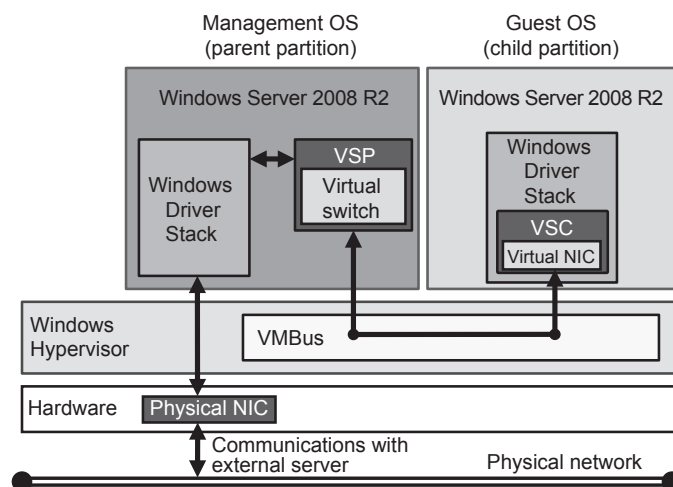


Figure 2
Architecture of virtual network.

entire memory contents temporarily. Since the system is running, the memory contents are still being updated even after the duplication. Nevertheless, Hyper-V performs a sequential update on the differential data as well by tracking and managing the modified portions. The operation of the VM is halted temporarily at a point at which the difference becomes sufficiently small. Differential information on the various items of context and memory that the VM retains are then transferred, and finally the transition is completed by switching ownership of the shared disk to the server at the transition destination. Note that live migration is done between single VMs on each server.

6. Cluster shared volume

A cluster shared volume (CSV) is a disk-sharing function of Hyper-V that is newly supported by the Windows Server 2008 R2 OS and that extends the MSFC function. The MSFC function manages the shared disk resources device by device or logical unit number (LUN) by LUN. Since these resources are controlled by the server that has become the owner, servers other than the owner cannot access the disk. This means that when all VHDs used by VMs are located in one shared disk resource, all the VMs must run on the server that owns the shared disk resource so that they can access their VHDs on the shared disk resource.

To enable live migration using a standard shared disk resource, therefore, it is necessary to assign a dedicated LUN to each VM, and the server at the transition destination must be switched to being the owner of the shared disk resource (the dedicated LUN) during the transition of the VM. As a result, with a live migration configuration in which a standard shared disk resource is used, time is required for switching over the disk in addition to the transition time for the VM.

The CSV function enables a number of servers to use the same LUN simultaneously,

thereby overcoming these problems.

6.1 CSV architecture

To implement simultaneous use of one LUN by multiple servers with integrity, it is generally necessary to use a dedicated cluster file system. The CSV function supports this by adding a new CSV filter layer onto NTFS, which is the standard Windows file system. The advantage of this architecture is that it is not a new file system. It makes use of the existing NTFS file system and its functions, which have proved themselves in use.

As shown in **Figure 3**, each LUN has an “owner node” within the CSV that manages and controls the LUN. When the VM on the owner node accesses the CSV, the operation is exactly the same as that of ordinary disk access, that is; the CSV filter driver is transparent.

If a VM operating on a non-owner node accesses the CSV [1] in Figure 3], on the other hand, the CSV filter driver acquires information about accessing the target file from the owner node [2]. The information acquired is used for direct access to the disk, without going through the NTFS file system driver [3].⁷⁾

In addition, the CSV filter driver communicates with the owner node when it is necessary to access the management information in the file system such as for expanding a VHD.

Thus, the CSV function maintains the integrity of the shared volumes by changing the behavior of the CSV filter driver depending on the situation.

6.2 Redirected I/O

The CSV function does not simply enable simultaneous accesses from a number of servers; it also provides high-level availability functions that enable access through other nodes if the path to the disk is broken. This is called “redirected I/O.” If the path between a node and a shared disk breaks, access to the CSV from the node would ordinarily be impossible. However,

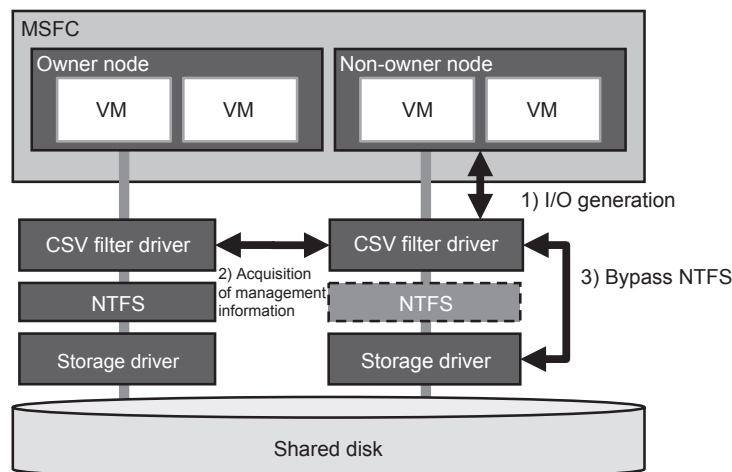


Figure 3
CSV disk access.

redirected I/O enables the node to re-access the CSV. The I/O access from the node to the CSV is redirected through the network to the owner node, and then the owner node performs the I/O access process instead of the node.

7. Hyper-V management by SCVMM

The Hyper-V virtualization platform is designed to be managed by the Hyper-V manager, a standard part of the OS, but it is not suitable for the collective management of a number of servers because it is intended for managing a single physical server.

Microsoft provides the System Center Virtual Machine Manager (SCVMM) product to solve this problem.⁸⁾ It improves the manageability of the Hyper-V environment and also enables collective management of different virtualization platforms such as the Virtual Server and VMware ESX. Here we discuss the SCVMM architecture and introduce the main functions provided by the SCVMM.

7.1 SCVMM architecture

The SCVMM is configured of the components shown in **Figure 4**, and all operations and commands are executed through Microsoft's

Windows PowerShell^{note)} commandlets. Windows PowerShell execution commands are generated by operating the SCVMM Administrator Console, which is a management tool, and processing passes to the SCVMM through the Windows Communication Foundation framework. Subsequently, commands are transferred through the SCVMM agent to the Hyper-V components and also to the library that stores resources to be deployed and provisioned.⁹⁾

7.2 Main functions of SCVMM

1) Setting of management access rights

The SCVMM is designed to enable easy setting of management access rights for each physical server or VM. The access authorities can be configured as shown in **Table 1**.

Note that a website dedicated to VM management, called a “self-service portal,” is provided for self-service users (users who operate only VMs). Self-service users access that site to operate their VMs.

2) Creation, deployment and provisioning of VMs

A number of VMs with the same settings

note) This is a command-line shell and scripting language developed by Microsoft. It is integrated with the .NET framework and supports automation of the management of Windows OS and applications.

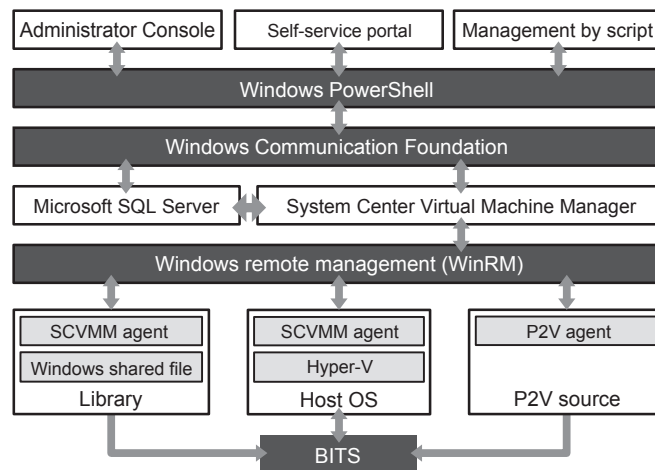


Figure 4
SCVMM configuration.

Table 1
Access authorities added by SCVMM.

Authority	Summary
Manager	Management of all servers
Proxy manager	Management of some servers
Self-service user	Operations that can be performed by users, such as activation or halting of VMs, can be specified in detail for each user

can be set up by using the VM template function of SCVMM.

Simply duplicating a virtual disk without using the template function, then using that copy as a basis for the creation of a new VM would cause problems because the computer security identifier (SID) would be duplicated, so the clone VMs would not be able to operate properly within the same network. Deployment and provisioning using the template function gives each new VM a unique SID, thereby avoiding duplication of the SID of a newly created VM with those of VMs at the destination. In addition, the effects of data transfer on other applications can be kept to a minimum by using the background intelligent transfer service (BITS) for data transfer. Furthermore, transfers can restart automatically, even if the network is restored after a network failure.

3) Migration from physical server to VM

A physical-to-virtual (P2V) process is provided to assist in the migration from an existing physical server to a virtual environment. With a server OS, the physical server can use online P2V to convert to a VM even when the physical server is running, provided it is Windows Server 2003 SP2 or later. With Windows 2000, only offline P2V can be used (when executed after the physical server has been halted).

When data is transferred by offline P2V, the transfer occurs after a Windows preinstallation environment (PE) image has been activated on the physical server at the conversion source. Because the Windows PE used by offline P2V is created on the basis of a Windows Vista kernel, drivers compatible with Windows Vista that run in the Windows PE might not exist on older physical servers. Nevertheless, even if the physical server does not have a device supported by a device driver in the Windows PE, an appropriate driver can be automatically transferred when the P2V process is executed. This automatic transfer simply requires that the driver be registered beforehand in the SCVMM.

8. Customer case study

Mitsui & Co., Ltd., is starting an initiative to consolidate the sequential virtualization of

1000 existing servers, aimed at implementing a highly flexible IT infrastructure that will be able to respond rapidly at reasonable cost to changes in business. The plan is to use the PRIMERGY RX300 S5 server for the PC servers in which Hyper-V will operate at the core of the IT infrastructure and to standardize applications, middleware, OS's, and hardware in rack units, resulting in a private cloud platform for the entire Mitsui & Co., Ltd., Group.

There are several reasons for Mitsui's selection of Fujitsu as their platform partner. Fujitsu has been at the cutting edge in working with Hyper-V and has acquired practical experience, technological strength, and problem-solving abilities from beta testing. In addition, the PRIMERGY server is sophisticated, is highly reliable, and has superior cost performance.¹⁰⁾

9. Conclusion

We have introduced the Hyper-V server virtualization function provided by Microsoft's Windows Server 2008 R2 OS. Hyper-V is continuing to evolve as a core technology supporting the next generation of the Windows server OS and Microsoft's private cloud. At Fujitsu, we plan to forge closer links with Microsoft and to support our customers' systems

and businesses as a vendor capable of providing one-stop support for all hardware and software.

References

- 1) Fujitsu: Solutions to IT Infrastructure Challenges by Hyper-V. (in Japanese). <http://primeserver.fujitsu.com/primergy/software/windows/os/wins2008/hv/>
- 2) Fujitsu: Information on design, configuration, and migration of Windows Server 2008/Windows Server 2008 R2. (in Japanese). <http://primeserver.fujitsu.com/primergy/software/windows/os/wins2008/>
- 3) R. Larson et al.: Windows Server 2008 Hyper-V Resource Kit. Microsoft Press, June 10, 2009.
- 4) M.E. Russinovich et al.: Windows Internals 5th Edition. Microsoft Press, June 17, 2009.
- 5) Virtual Hard Disk Image Format Specification. October 11, 2006 - Version 1.0 Microsoft.
- 6) A. Dabagh: Improving Networking Performance for Hyper-V Virtual Machines. WinHEC 2008, November 2008.
- 7) J. Mastro et al.: Windows Virtualization and Cluster Shared Volumes. WinHEC 2008, November 2008.
- 8) System Center Virtual Machine Manager 2008 and Virtual Machine Manager 2008 R2. <http://technet.microsoft.com/en-us/library/cc917964.aspx>
- 9) Microsoft: System Center Virtual Machine Manager 2008 Practical Guide v1.0. (in Japanese). <http://technet.microsoft.com/ja-jp/systemcenter/vmm/dd433024.aspx>
- 10) Fujitsu: Mitsui & Co., Ltd., Introduction Case Study. (in Japanese). <http://primeserver.fujitsu.com/virtual/casestudies/mitsui/>



Yutaka Haga
Fujitsu Ltd.

Mr. Haga is engaged in design and development utilizing the Windows Server OS.



Masayuki Jibu
Fujitsu Ltd.

Mr. Jibu is engaged in design and development utilizing the Windows Server OS.



Kazuhide Imaeda
Fujitsu Ltd.

Mr. Imaeda is engaged in design and development utilizing the Windows Server OS.