Virtualization Technology of UNIX Server SPARC Enterprise Series for Reduced Total Cost of Ownership

● Yuuji Konno ● Hiromi Fukumura ● Hiroshi Murakami

Fujitsu has achieved high levels of quality and performance in its UNIX servers developed for mission-critical systems by using a variety of mainframe technologies developed over the years. It has pursued consolidation functions based on virtualization technology from an early stage to reduce both the initial investment and the total cost of ownership (TCO). With greater server processing power and functionality made possible by technical advances, these efforts have paid off in finer granularity of virtual resources and higher levels of consolidation that reduce TCO even more. Fujitsu has also worked to optimize initial investments through functions like central processing unit (CPU) mixing and to reduce management costs by unifying server management. This paper describes how Fujitsu is reducing TCO over the life cycle of an information and communications technology (ICT) platform while protecting performing assets and ensuring system stability through its SPARC Enterprise series of UNIX servers.

1. Introduction

Customer requirements for information and communications technology (ICT) infrastructures to support their business are mounting, and servers, which lie at the core of ICT infrastructures, are becoming key elements in meeting these requirements. Every year, we see more customers who would like to reduce their total cost of ownership (TCO) by making more effective use of resources throughout their systems. In response to this demand, Fujitsu has developed and released the UNIX server SPARC Enterprise M series (M9000, M8000, M5000, M4000, and M3000) supporting mission-critical systems and the UNIX server SPARC Enterprise T series (T5440, T5240, T5220, T5140, and T5120) for high-throughput computing. Both of these series inherit sophisticated technologies developed for Fujitsu's mainframe systems.

The SPARC Enterprise M series consists of highly versatile UNIX servers that can be applied to a broad range of enterprise business. The systems use the SPARC64 VI and SPARC64 VII SPARC processors developed by Fujitsu.

The SPARC Enterprise T series, on the other hand, is a family of UNIX servers applicable for Web front-end operations and application businesses. The systems use the UltraSPARC T2 and UltraSPARC T2 Plus high-throughput processors developed by Sun Microsystems.

The SPARC Enterprise series (hereinafter, SPARC Enterprise) uses Solaris as an operating system (OS) conforming to the UNIX OS international standard supporting many industry specifications. SPARC Solaris maintains binary compatibility between OS versions and guarantees the continuing usability of Solaris assets that the customer has already invested in.

This paper introduces Fujitsu's efforts to reduce TCO including the customer's initial investments through "virtualization technology," a key element of SPARC Enterprise.

2. Problems with server implementation and operation

Server-related problems are broadly divided into those occurring at the time of server implementation and those occurring during server operation.

1) Server implementation problems

Migration to a new system requires some time for updating middleware and testing the operation of customer applications, and many migrations do not go as smoothly as desired. Increasing the time spent for testing generally increases the implementation cost, so sufficient testing should be done as quickly as possible. Updating applications and middleware to run on the new system also incurs an additional expense.

2) Server operation problems

Increasing the number of servers to handle an increase in business processing may lead to insufficient installation space and a rise in operating costs due to higher electricity bills. Adding servers can also make operation management more complicated, which can lead to higher management costs.

The following sections describe Fujitsu's approach to TCO reduction including the cost of implementing SPARC Enterprise with the aim of solving the above problems.

3. Addressing server implementation problems

SPARC Enterprise can reduce the cost of migrating to a new system through the use of virtualization technology in the form of Solaris 8/9 Containers. It can also reduce additional investments at the time of business expansion and server integration by making more effective use of hardware resources through functions like mixed mounting of new and existing central processing units (CPUs) and cabinet expansion. These measures are described in more detail below.

3.1 Solaris 8/9 Containers

Solaris 8/9 Containers is a virtualization function that enables Solaris 8 or Solaris 9 and its application environment to run on a SPARC Enterprise system with Solaris 10 running on it. It allows users to construct a highly reliable system with robust security and even higher application performance.

While operating under Solaris 10—the latest OS—is desirable to extract maximum performance from the system, migrating to Solaris 10 requires updating the middleware and allocating time for testing the operation of existing applications. Extending the testing period means increasing personnel expenses and other costs. Moreover, it is often impossible to secure a sufficient testing period from the start owing, for example, to the large number of applications. Therefore, there is a need for a migration process that can be performed quickly at low cost. Migration to Solaris 10 using Solaris 8/9 Containers as a means of quick low-cost migration is described below.

The use of Solaris 8/9 Containers on SPARC Enterprise is outlined in **Figure 1**. First, customer assets on existing server models (PRIMEPOWER and S series) are moved in their present form to a new SPARC Enterprise system by using Solaris 8/9 Containers. Next, while the system is operated in the originally constructed configuration, a test environment for the customer's assets on Solaris 10 is constructed and implemented in parallel on the same server (SPARC Enterprise). In other words, Solaris 8/9 Containers enables business operations to continue while migration to Solaris 10 and migration testing are simultaneously performed on a single server (one partition). And then, upon completion of testing, the system can be launched on Solaris 10 in a true operating environment and the old operating environment can be deleted.

The Solaris 8/9 Containers function therefore enables a customer system constructed on Solaris 8 or Solaris 9 to run on SPARC Enterprise

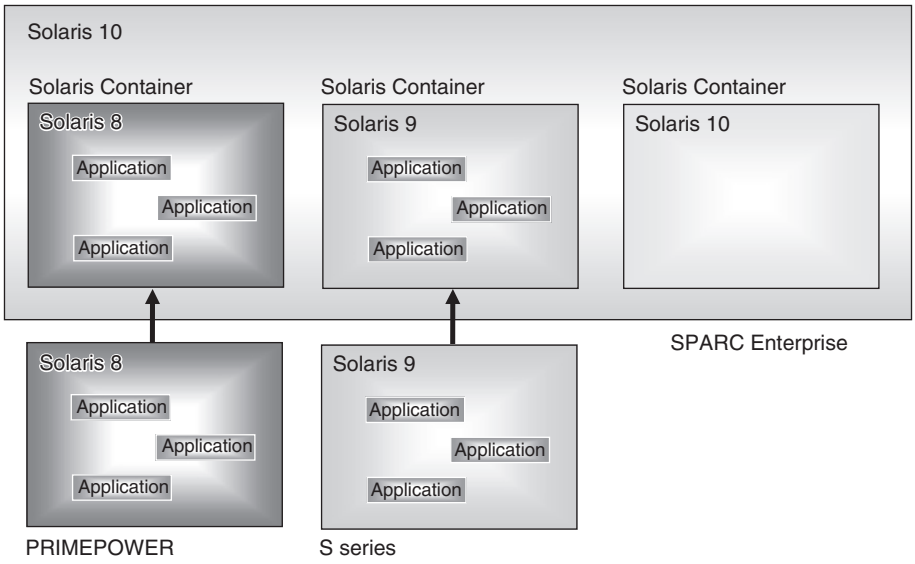

Figure 1
Solaris 8/9 Containers on SPARC Enterprise.

without any middleware or applications having to be modified. Thus, even when sufficient testing time for migration to a new server cannot be secured, migration to the new Solaris environment can be performed in stages in accordance with the operating conditions and migration circumstances of the customer's system. Even when migration to a new Solaris OS is not being undertaken yet, the use of Solaris 8/9 Containers can increase business processing speeds because existing business applications will be running on the latest hardware.

In short, as Solaris 8/9 Containers enables existing middleware and applications to be used without modification on Solaris 10, customer assets can be migrated to a new server rapidly at low cost.

3.2 Mixed mounting of new and existing CPUs

SPARC Enterprise supports mixed CPUs in the same chassis or partition, which means that a customer can add new CPUs (SPARC64 VII) effectively while using existing CPUs (SPARC64 VI) in the previously deployed chassis. CPU frequency mixing is also supported so that new CPUs can be added and performance can be improved without existing hardware resources having to be thrown away. For example, if new CPUs have been added to accommodate business expansion in a system that was initially configured with a few CPUs, the ability to operate the system with a mixture of old and new CPUs helps to keep down costs associated with initial investments. SPARC Enterprise contributes to the long-term protection of customer investments through this technology, enabling flexible system construction.

3.3 Cabinet expansion

In the case of the SPARC Enterprise M9000 server (32-CPU configuration), which consists of CPUs, CPU memory units (CMUs)[note 1], and input/output (I/O) units (IOUs) equipped with Peripheral Component Interconnect Express (PCIe) cards such as network cards, a system having a maximum of 8 CMUs and 8 IOUs can be constructed. This equals 32 CPUs, a memory capacity of 2 TB, and 64 PCIe short cards. A partitioning function enables the system to be

note 1) A CMU is a system board containing CPUs and memory.

divided into a maximum of 24 partitions.

The system can be expanded to a maximum of 16 CMUs and 16 IOUs (64-CPU configuration) by connecting to an M9000 expansion cabinet via a crossbar cable. This configuration accommodates up to 64 CPUs, up to 4 TB of memory, and 128 PCIe short cards. The system can also be divided into a maximum of 24 partitions.

The use of a crossbar cable to connect to the same type of chassis enables SPARC Enterprise to provide high-speed processing of transactions, generating an even bigger load on the system, without major changes having to be made to that system. This feature helps to reduce implementation costs.

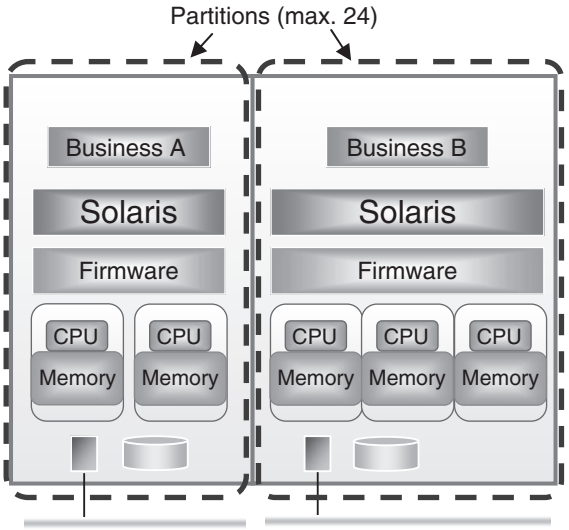

Figure 2
Hardware partitioning.

4. Addressing server operation problems

SPARC Enterprise supports three virtualization functions: hardware partitioning, Logical Domains (LDoms), and Solaris Containers. These functions enable multiple servers to be integrated on a single server (server consolidation) and reduce the required server installation spaces and power consumption. Moreover, a CMU hot-swap function can be used during system operation to help minimize planned system downtime when CPUs are added or components are switched. Furthermore, the eXtended System Control Facility (XSCF) enables uniform management of multiple partitions, thereby making a dedicated management server unnecessary and keeping management costs down. This section describes specific technologies for reducing TCO in server operation.

4.1 Server consolidation

The three virtualization functions that enable SPARC Enterprise to achieve server consolidation are described below.

1) Hardware partitioning

The SPARC Enterprise hardware-partitioning function is outlined in **Figure 2**. The SPARC Enterprise M series has a partitioning function for constructing multiple systems by dividing a single server into multiple regions (partitions) and running an OS and applications in each partition. A minimum of one CPU is required to create a partition and a maximum of 24 partitions can be configured. Each partition is physically independent, so the failure in one partition does not influence any other partition, which enables operations to continue. Partition creation is controlled by the XSCF function in such a way that the performance of each partition will not be affected.

In the past, multiple servers could share hardware through the use of middleware. In contrast, the SPARC Enterprise partitioning function enables reserved resources on a single server to be shared by more than one partition. Compared with the use of dedicated middleware, this partitioning function eliminates the need to change middleware settings as would normally be required for moving CPUs, memory, and other resources. Another of its features is small units for resource addition (upgrading when using middleware is usually performed in units of systems). Therefore, the partitioning function enables highly flexible system construction.

In the above approach, the SPARC

Enterprise hardware partitioning function allows one to use only one server to achieve business operations that in the past required several small-scale servers.

2) LDoms

LDoms is a technique for consolidation at the firmware level (**Figure 3**). The SPARC Enterprise T series supports LDoms for constructing multiple systems by configuring multiple virtual hardware environments (domains) at the firmware level and running an OS and applications in each domain. LDoms consists of a "control domain" and "guest domains." The control domain is in charge of creating and managing all domains and provides guest domains with virtual services and virtual I/O through the Logical Domains Manager. A guest domain runs business applications using virtual services and virtual I/O under the management of the control domain. It is a completely independent Solaris OS environment that can be booted and shut down without affecting other domains. A domain can add or delete CPU resources and I/O without having to be shut down. Since LDoms manages hardware resources at the firmware level, the CPU, I/O, and memory resources can be divided into virtual units that are smaller than the physical units. Up to 128 domains can be created, which means that business-application consolidation by the finer allocation of resources achieves a higher density than that achieved by hardware partitioning.

3) Solaris Containers

Solaris Containers is a software-level consolidation technique (**Figure 4**). It is a virtual technique provided by Solaris that can be used in either the SPARC Enterprise M or T series of servers. It enables the OS to be divided in a virtual manner so that resources can be used flexibly according to the needs of each business application.

Solaris Containers consists of the Solaris Zone and Solaris Resource Manager functions. Solaris Zone is a software partitioning function that divides a single OS space into multiple virtual partitions to make it appear as if multiple

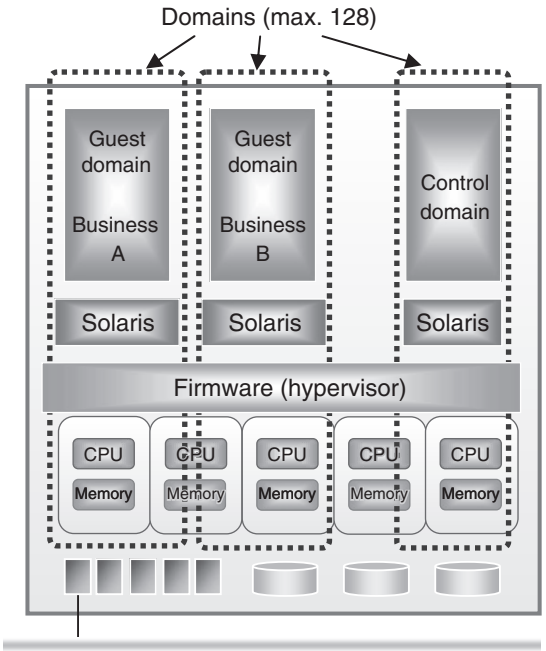

Figure 3
LDoms (firmware level).

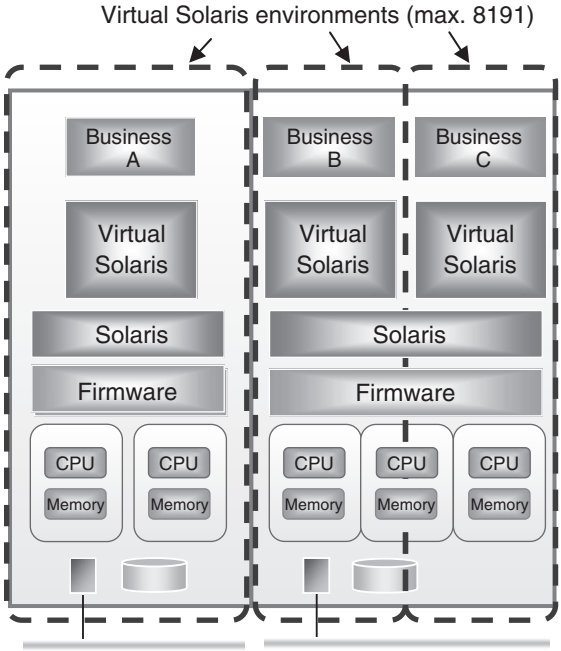

Figure 4
Solaris Containers (software level).

operating systems are running. A maximum of 8191 virtual Solaris environments can be created. The Solaris Resource Manager function allocates hardware resources such as CPUs and memory in a flexible manner across virtual Solaris environments. For example, it can configure multiple virtual environments even on a single-CPU server to make more efficient use of hardware resources. It can also prioritize the allocation of resources to business applications that have a high usage frequency.

The hardware partitioning function and Solaris Containers can be combined to construct a highly flexible and available system.

4.2 CMU hot-swapping

The CMU hot-swap function is used to upgrade hardware when servers are being integrated while the system is operating. In the SPARC Enterprise M9000 and M8000 series, hardware resources such as CPUs and memory can be added or exchanged during system operations without shutting down partitions in conjunction with the dynamic reconfiguration[note 2)] function. Furthermore, to meet the demands for business expansion or new business applications, this hot-swap function can be used to flexibly add or allocate hardware resources without influencing business operations running in other partitions. Thus, by enabling components to be exchanged without a system shutdown, the CMU hot-swap function helps to reduce operating costs.

4.3 XSCF system monitoring mechanism

When a system is being constructed, it is a general practice to construct separate servers for managing multiple domains and for configuring the system. SPARC Enterprise, however, enables virtual domains divided into multiple partitions to be uniformly managed by XSCF, which makes a dedicated management server unnecessary and simplifies the management of each domain.

The XSCF function provides a system monitoring mechanism that runs on a service processor independent of the system CPUs on the main unit. It monitors the state of various types of hardware (CPUs, memory, disks, etc.) as well as cooling fan speeds and equipment temperatures. The XSCF function continues to monitor and control the main unit even if the OS goes down. The gathered information is logged and stored, so that errors can be isolated and the causes will be analyzed quickly even in the case where the OS cannot be booted. The XSCF function has a redundant configuration in the SPARC Enterprise M9000 and M8000 series to support business operations that require exceptionally high reliability.

In addition to command-line operation, XSCF can be accessed from a Web browser with communications secured by Secure Shell (SSH) or Secure Sockets Layer (SSL). Since the state of the main unit can be easily and accurately determined from XSCF, the work load of system managers can be lightened and server operating costs will be reduced.

5. TCO reduction effect in server implementation and operation

The SPARC Enterprise functions described above are expected to produce the following effects.

1) The use of Solaris 8/9 Containers will reduce the migration period to new servers, the expense incurred in upgrading middleware in a migration, and personnel costs associated with testing.
2) Cabinet-expansion and CPU-mixing technologies will enable the customer's existing hardware resources to be used more effectively and will optimize and reduce the cost of adding new hardware.
3) Server integration by virtualization will

note 2) Dynamic reconfiguration is technology for moving hardware resources like CPUs among partitions during partitioning operations.

reduce operating costs related to power consumption, installation space, etc.

4) Unified management of multiple partitions by XSCF will lighten the work load on managers.

6. Conclusion

This paper introduced Fujitsu's measures for reducing TCO through virtualization technology in its SPARC Enterprise. Virtualization technology enables multiple servers to be consolidated in one SPARC Enterprise system, and CPU mixing and other functions make for more effective use of hardware resources, enabling initial investments to be protected and operation and management costs to be reduced.

Looking forward, Fujitsu will continue to assist the customer in expanding its business by pursuing finer granularity of virtual resources and higher levels of consolidation, achieving further reductions in TCO, protecting the customer's ICT investments, and providing even better systems.

Yuuji Konno
Fujitsu Ltd.
Mr. Konno is engaged in the testing of SPARC Enterprise systems.

Hiromi Fukumura
Fujitsu Ltd.
Ms. Fukumura is engaged in the study of SPARC Enterprise system specifications.

Hiroshi Murakami
Fujitsu Ltd.
Mr. Murakami has worked in ASIC development and testing for SPARC Enterprise systems and is currently engaged in the testing of Cloud environments in Cloud infrastructure solution center.