

# Building a High Availability System on Fujitsu SPARC M12 and Fujitsu M10/SPARC M10 Servers (Overview)

September 2017

Rev 3.0

Fujitsu LIMITED

## Introduction

1. Model Case of HA System
2. Example of HA System
3. Behavior upon Hardware Failure
4. Hot replace Procedure
5. Important Notes

Reference: Key Fujitsu SPARC M12 and Fujitsu M10 Features for High Availability

## ■ Preface

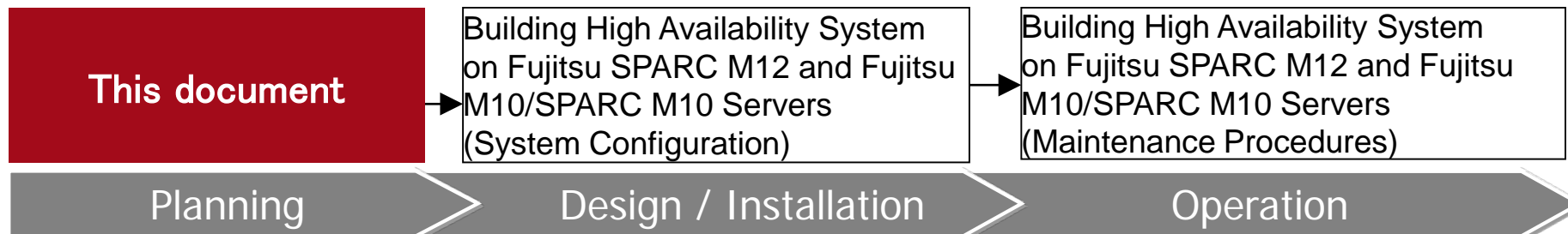
- This document is intended to introduce features, advantages and guidelines for a Building Block High Availability (BB HA) system.

## ■ Audience

- System Administrators who are planning to build an HA system on Fujitsu SPARC M12 and Fujitsu M10 Servers.
- System Administrators with advanced knowledge of Oracle Solaris/Oracle VM Server for SPARC.

## ■ Notes

- In this document, Oracle Solaris may be referred to as Solaris, and Oracle VM Server for SPARC as Oracle VM or OVM.
- This document provides an overview of what a BB HA system can do.  
For more information, see :
  - Building a High Availability System on Fujitsu SPARC M12 and Fujitsu M10/SPARC M10 Servers (System Configuration)
  - Building a High Availability System on Fujitsu SPARC M12 and Fujitsu M10/SPARC M10 Servers (Maintenance Procedures).



# Building Block High Availability (BB HA)

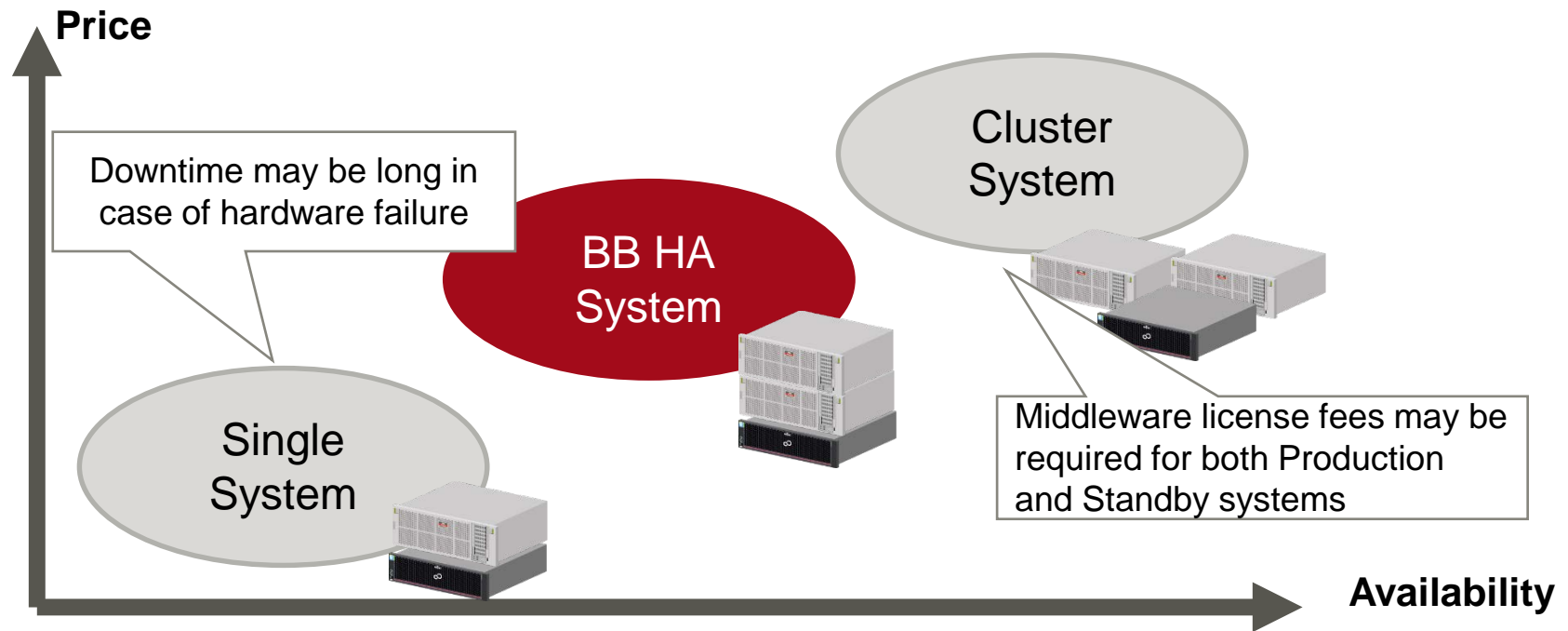
## Key Points of a Building Block High Availability system (BB HA system)

Self-recovery from hardware failure and resume work  
Hot replace of faulty parts

Increase  
Availability

Eliminate middleware license fees for a standby system

Eliminate  
Cost


















Fujitsu SPARC M12-2S and Fujitsu M10-4S provides  
higher availability with lower cost

# Selection Criteria for System Configuration

## ■ Increase availability and eliminate costs

- Cluster system has better availability than BB HA but costs more.
- If up to an hour of downtime is acceptable, BB HA is the best choice.

		Single system	BB HA system	Cluster system
Availability	Monitored resources	 N/A	 Hardware failure	 Hardware failure OS failure, Application failure
	Downtime on failure	 N/A (Recover manually)	 Minutes to an hour	 Seconds to minutes
	Downtime on repair	 Cold replace	 Minutes (Hot replace by PPAR Dynamic Reconfiguration (*))	 Minutes (Hot replace by switching server)
Cost	Middleware	 Production system only	 Production system only	 Production and standby system
	Cluster software	 N/A	 N/A	 Required

(\*) PPAR Dynamic Reconfiguration (PPAR DR): See P. 35 for more details.

# High Availability on Fujitsu SPARC M12-2S and Fujitsu M10-4S

■ Fujitsu SPARC M12 and Fujitsu M10 features enable you to configure a high availability system.

## ■ Building Blocks

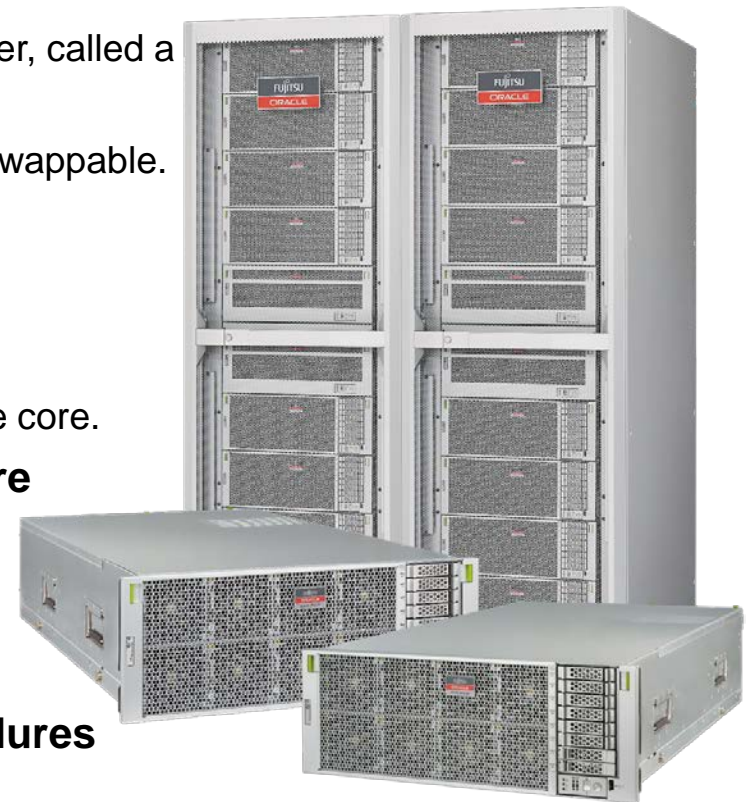
- Combine Building Blocks (BB) to create a scaled-up server, called a Physical Partition (PPAR).
- Each Building Block is dynamically expandable and hot swappable.
  - **BBs can be configured dynamically.**

## ■ CPU Core Activation

- Upgrade CPU resources as needed (pay-as-you-grow).
- Faulty CPU cores are automatically replaced with a spare core.
  - **Auto recovery in case of hardware failure**

## ■ PPAR Dynamic Reconfiguration

- Expand resources without business interruptions.
- Reduce downtime during hardware replacements.
  - **Minimize downtime during hardware failures**




This document describes a configuration that minimizes system downtime by combining Fujitsu SPARC M12-2S and Fujitsu M10-4S Building Blocks and PPAR Dynamic Reconfiguration (PPAR DR).

# 1. Model Case of HA system



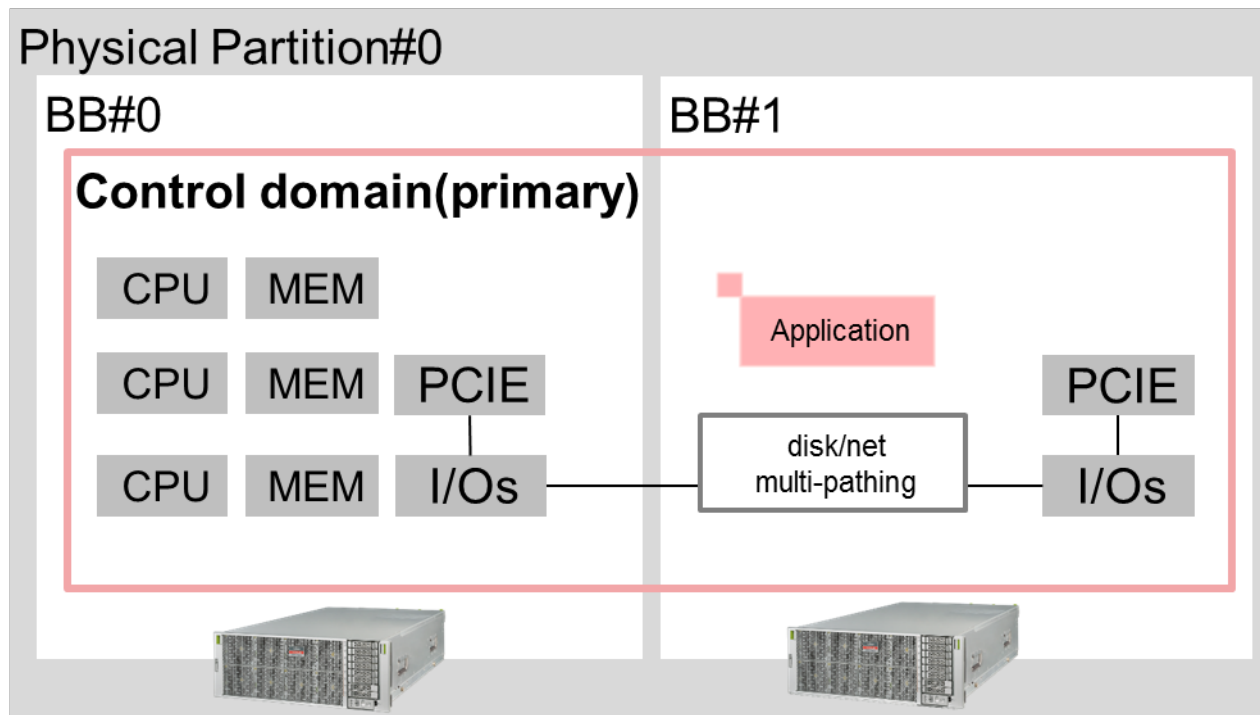
# Model Case of HA system

Type	Overview	Domain configuration	Skill to configure
<b>Traditional</b>	Running only one domain with a PPAR.	Control domain only. Solaris Zones can be configured.	<b>Easy</b>
<b>Server consolidation</b>	Configure multiple root domains within a PPAR using OVM and consolidate multiple servers into one system. Assigning Direct I/O to domains offers better I/O performance.	Control domain and root domains. (Up to 4 root domains with a 2 x Fujitsu SPARC M12-2S and Fujitsu M10-4S configuration)	 <b>Hard</b>
<b>High consolidation</b>	Configure multiple guest domains and root domains within a PPAR using OVM. More domains can be consolidated than "Server consolidation" using virtual I/O.	A or B is available.  A) Control domain and multiple guest domains.  B) Two root domains and multiple guest domains.	

# Traditional Type

## ■ Features of a traditional HA system

- CPU Auto Replacement will run on CPU core failure to maintain system performance.
- Multi-pathing I/O for redundancy.
- BB is hot swappable in the case of hardware failure.
- **System runs with a single domain.**

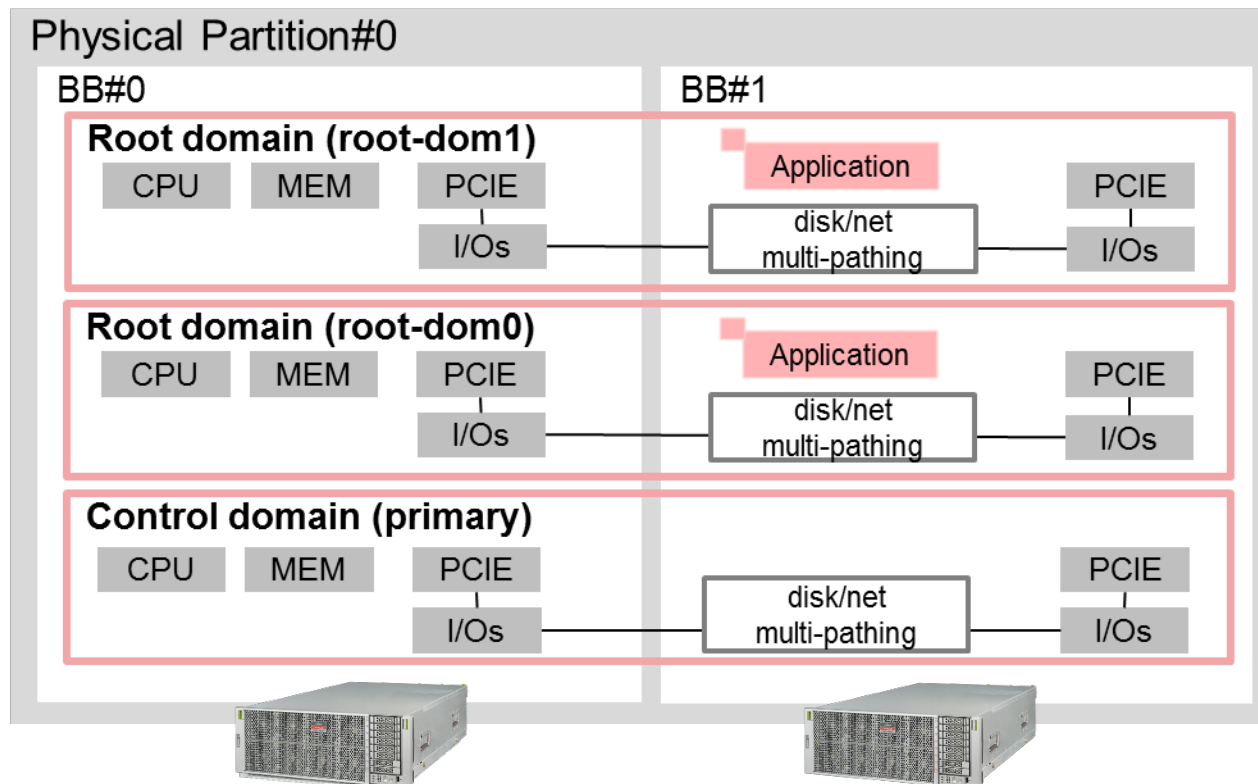


# Server Consolidation Type

## ■ Features of a server consolidation HA system

### - Consolidate multiple business applications into a single PPAR

- CPU Auto Replacement will run on CPU core failure to maintain system performance.
- Multi-pathing I/O for redundancy.
- BB is hot swappable in the case of hardware failure.
- Runs business applications on root domains.
- **Consolidate multiple business applications into a single PPAR (up to 4 root domains)**

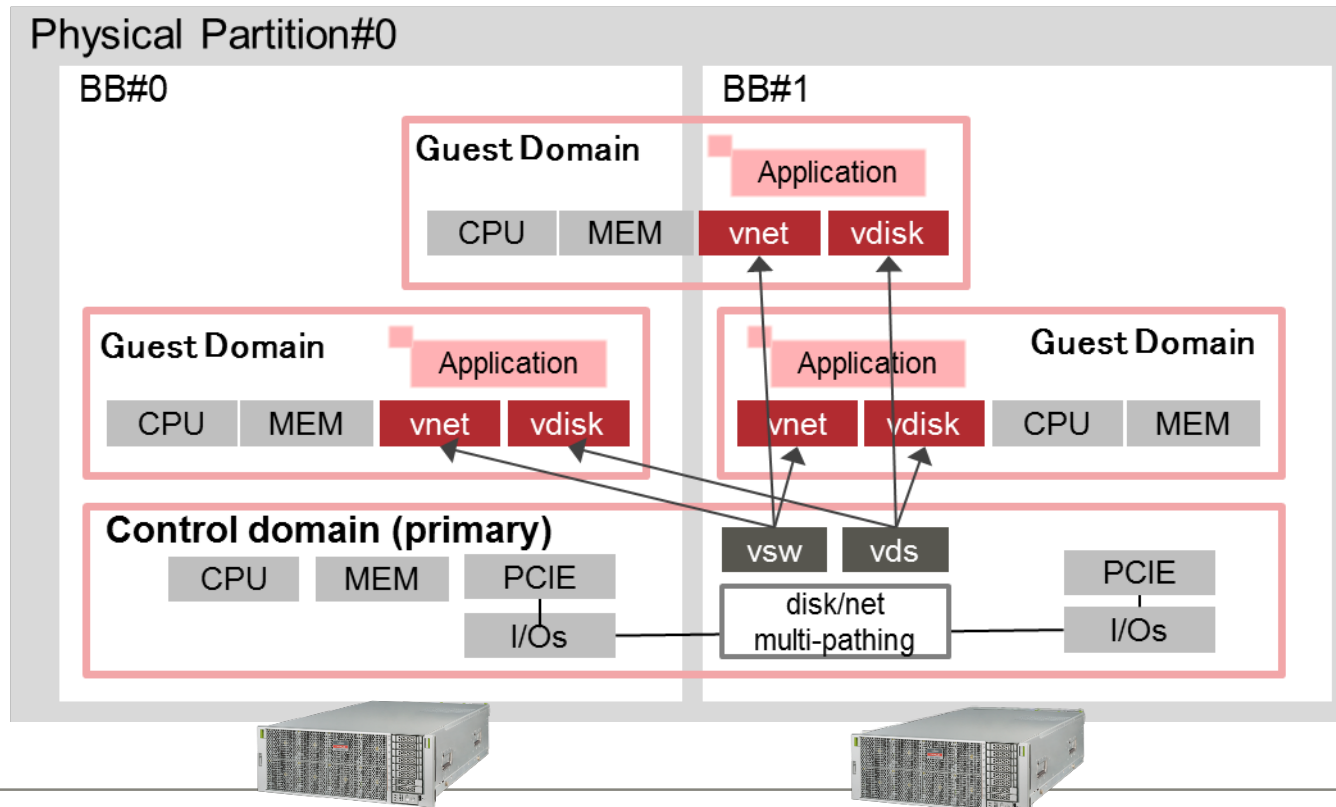


# High Consolidation Type A

## ■ Features of a high consolidation HA system

### - By using virtual I/O, more domains can be consolidated

- CPU Auto Replacement will run on CPU core failure to maintain system performance.
- Multi-pathing I/O for redundancy.
- BB is hot swappable in the case of hardware failure.
- Runs business applications on guest domains.
- **Consolidate multiple business applications into a single PPAR (more than 4 domains)**

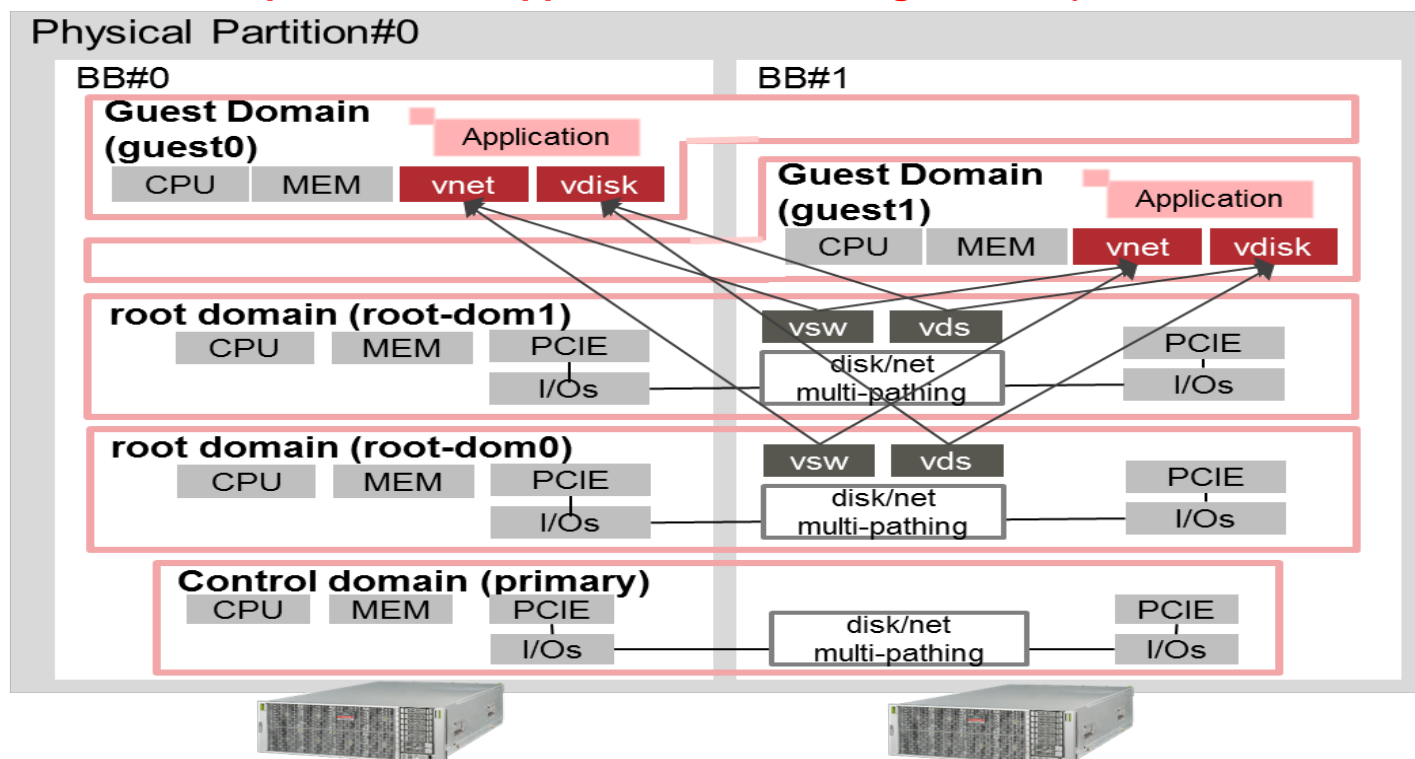


# High Consolidation Type B

## ■ Features of a high consolidation HA system

### - By using virtual I/O, more domains can be consolidated

- CPU Auto Replacement will run on CPU core failure to maintain system performance.
- Multi-pathing I/O for redundancy.
- Two root domains for roll update.
- BB is hot swappable in the case of hardware failure.
- Runs business applications on guest domains.
- **Consolidate multiple business applications into a single PPAR (more than 4 domains)**



## ■ Hardware requirements

- Servers : 2 BBs (2 Fujitsu SPARC M12-2S or 2 Fujitsu M10-4S) in a physical partition.
- I/O cards : Redundant using a pair of cards.
- Memories :
  1. Each BB must have the same physical memory configuration, which means the same capacity DIMMs must be installed in the same position across all BBs.
  2. Each memory group (16 DIMM slots associated to a CPU socket) must satisfy the following capacity limitation:
$$\text{CPU\#0 Group A} \leq \text{CPU\#0 Group B} \leq \text{CPU\#1 Group A} \leq \dots \leq \text{CPU\#3 Group B}$$

## ■ Software requirements for Fujitsu SPARC M12

- Firmware version : XCP 3021 or later
- OS version on control domain : Oracle Solaris 11.3 SRU11.3.17.5.0 or later \*
- OS version on root domain : Oracle Solaris 11.3 SRU11.3.17.5.0 or later
- OS version on guest domain : Oracle Solaris 11.3 SRU11.3.17.5.0 or later  
Oracle Solaris 10 1/13 or later \*

\* For High Consolidation type B, control domain is required Oracle Solaris 11.3 SRU11.3.23.5.0 or later and guest domains are required Oracle Solaris 11.3 SRU11.3.17.5.0 or later.

# Prerequisites (For all HA system types)

## ■ Software requirements for Fujitsu M10\*

- Firmware version : XCP 2240 or later
- OS version on control domain : Oracle Solaris 11.2 SRU11.2.8.4.0 or later
- OS version on root domain : Oracle Solaris 11.2 SRU11.2.8.4.0 or later
- OS version on guest domain : Oracle Solaris 11.1 SRU1.4 or later  
Oracle Solaris 10 1/13 or later

\* High Consolidation Type B is not qualified at this time, but should work; if this configuration is of interest, contact M12\_force@us.fujitsu.com.

## ■ Note for Fujitsu SPARC M12 and Fujitsu M10

It is strongly recommended that the latest version software and firmware are applied to the system to avoid the known issues. Please refer to the following the known issues for details.

- Building a High Availability System on Fujitsu SPARC M12 and Fujitsu M10/SPARC M10 Servers (System Configuration)
- Building a High Availability System on Fujitsu SPARC M12 and Fujitsu M10/SPARC M10 Servers (Maintenance Procedure)

To activate the PPAR DR parameter, PPAR must be reset after changing the parameter.  
Please see the setpparmode section in XSCF reference manual for more information.  
To enable Recovery Mode, the OVM variable must be enabled.  
Please see the Oracle VM for SPARC Administration Guide for more information.

See Fujitsu SPARC M12 and Fujitsu M10-4S manuals and product notes for the latest information.

<http://www.fujitsu.com/global/products/computing/servers/unix/sparc/downloads/manuals/m12-2s/>  
<http://www.fujitsu.com/global/products/computing/servers/unix/sparc/downloads/manuals/m10-4s/>

# Requirements (For all HA system types)

## ■ Physical Partitions

- A physical partition has 2BBs(2 Fujitsu SPARC M12-2S or 2 Fujitsu M10-4S), to recover the business in the case of hardware failure.

## ■ CPUs

- Half of CPU cores must be activated, the remaining half is kept for pool.
  - This enables CPU Auto Replacement on any failure.
  - This will prevent CPU performance degradation during Hot replace to maintain the number of running CPUs.

## ■ Memory

- About half of the memory can be assigned to domains, the rest must be reserved as a pool. Hypervisor uses system memory in the following table, (total installed memory / 2 – hypervisor memory) can be assigned to domains.

This will prevent lack of memory during Hot replace to maintain memory capacity.

	Fujitsu SPARC M12-2S	Fujitsu M10-4S
PPAR DR Enabled	4GB	3.25GB
PPAR DR Disabled	4GB	3GB

## ■ I/O

- I/O on each domain must be redundant across BBs.
  - This will keep providing active I/Os during Hot replace .

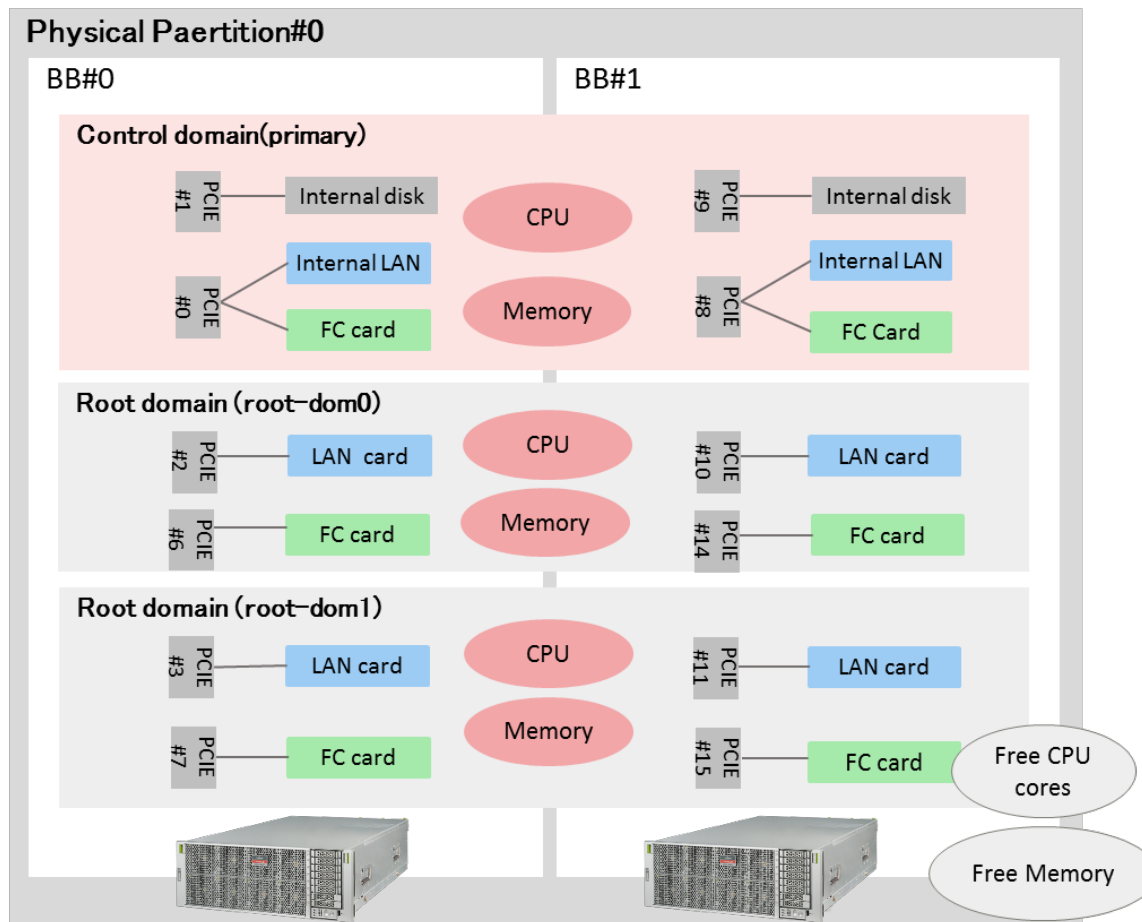


## 2. Example HA system [Server Consolidation Type]

# Example of HA system [Server Consolidation Type]

## ■ Consolidate multiple business applications into single PPAR.

- To maintain CPU/Memory resources during Live Service, CPU/Memory in 1BB should be reserved.
- Multi-pathing I/O is used for redundancy across BBs in each domain.



## ■ Control domain

- CPU: 1 or more cores / Memory: 8GB or more

[Note] Please estimate CPUs and memory according to the actual system environment.

## ■ Virtual Services

- Virtual Console Concentrator(vcc) is enabled on the control domain.

## ■ Resource assignment

- Do not specify core id to assign vCPUs.
- Do not specify Physical Address to assign memory.
- PCIe Buses that own internal disk(s) are assigned to the control domain.

## ■ System volume

- Control domain uses SAN disk for the system volume.
- Using FC cards on each BB to multi-path the SAN disk.

## ■ NTP client

- Performing PPAR DR may suspend / resume the OS. Since suspend / resume may cause a time delay, so the NTP client should be configured to adjust time automatically.

## ■ Root domain

- CPU: 1 or more cores / Memory: 4GB or more

[Note] Please estimate CPUs and memory according to the actual system environment.

## ■ Resource assignment

- Do not specify core id to assign vCPUs.
- Do not specify Physical Address to assign memory.
- PCIe Buses are assigned to root domains to configure I/O multi-pathing.

## ■ System volume

- Root domain uses SAN disk for the system volume.
- Using FC cards on each BB to multi-path the SAN disk.

## ■ Applications

- Business applications are installed in root domains.

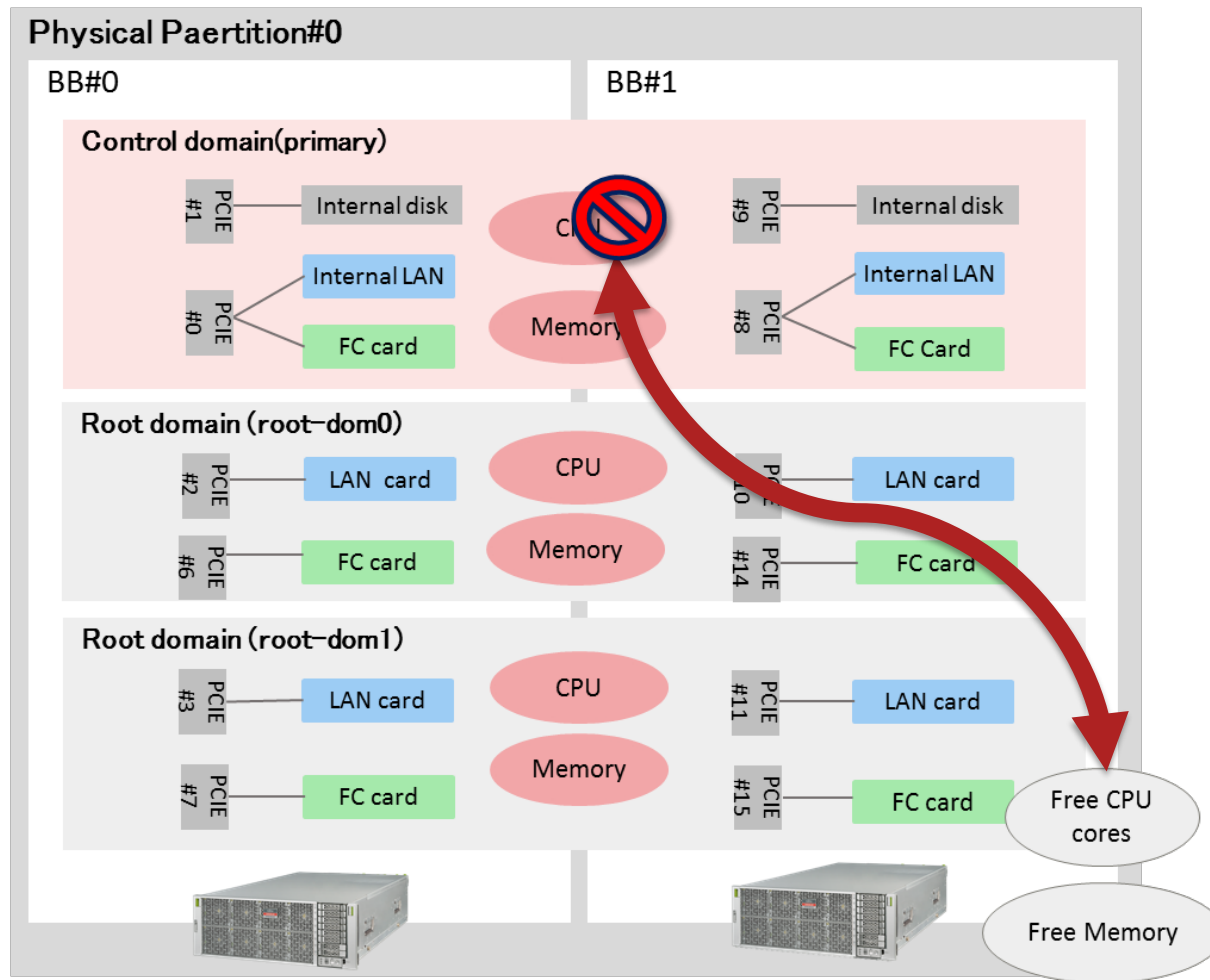
## ■ NTP client

- Performing PPAR DR may suspend / resume the OS. Since suspend / resume may cause a time delay, so the NTP client should be configured to adjust time automatically.

### 3. Behavior upon Hardware Failure [Server Consolidation Type]

# Behavior upon CPU Core Failure

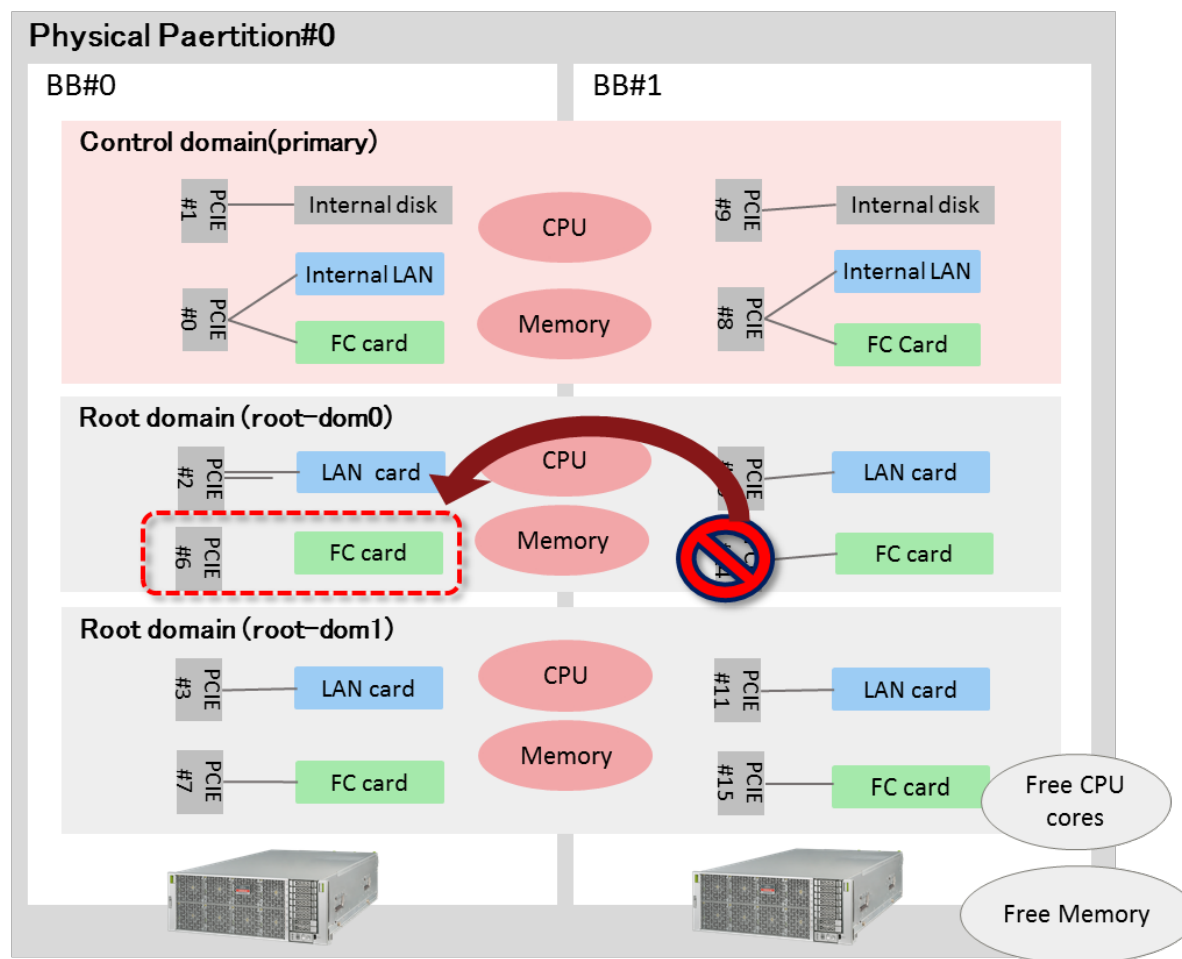
Replaces faulty CPU cores with deactivated CPU cores automatically. No downtime is required to replace CPU cores. (CPU core failures may cause OS panic, so there may be 10 to 15 minutes of downtime to reboot the OS.)



# Behavior upon PCIe Bus Failure

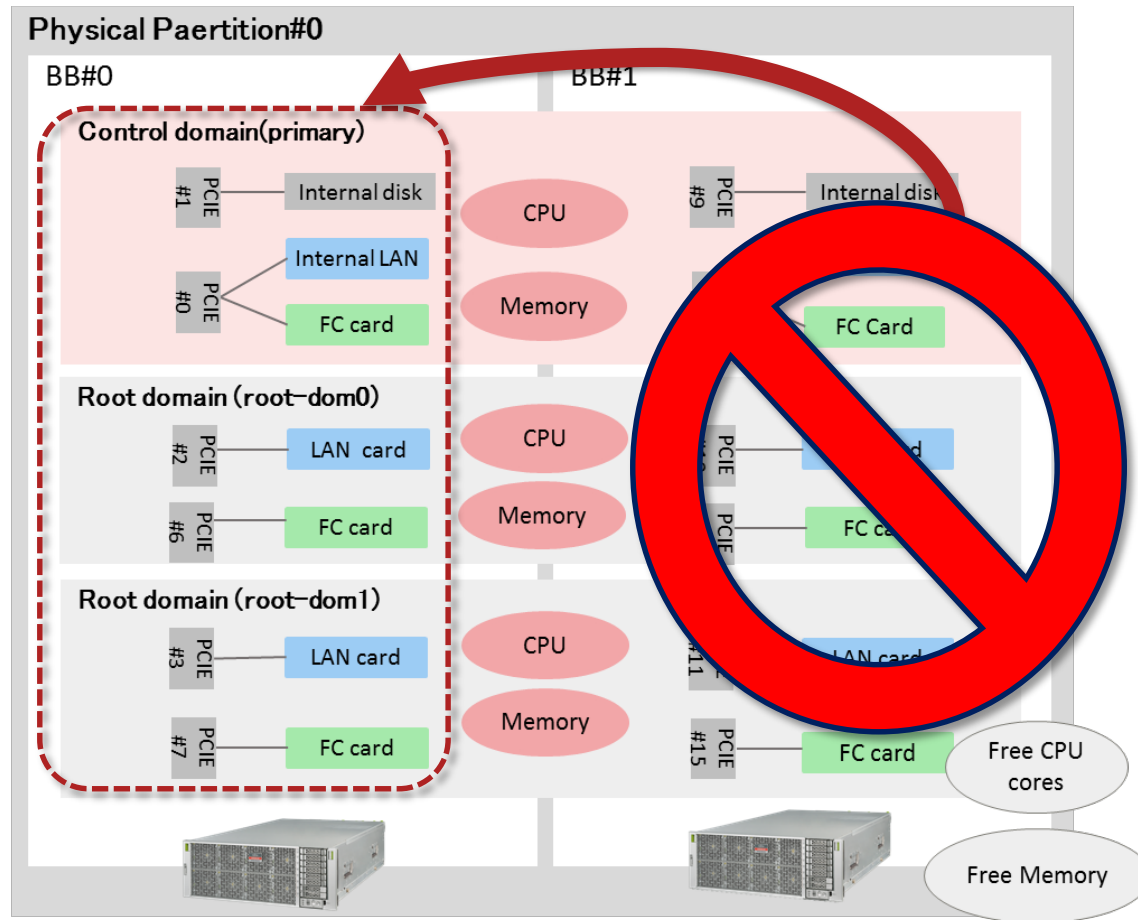
When a PCIe Bus failure happens, the OS may panic or a PPAR reset may occur.

After Rebooting OS/PPAR, faulty I/O resources are removed from the configuration and the system starts running with the remaining I/O resources. (Downtime: OS reboot 10 to 15 minutes / PPAR reset: 1 hour)



# Behavior upon BB Failure

In the case of one entire BB failure, after PPAR reset, OVM recovery mode reconfigures domains by removing faulty I/O resources from the configuration and re-assigning CPU and Memory. The number of CPUs and amount of memory are kept constant throughout the failure. (Downtime: 1 hour)





## 4. Hot replace Procedures [Server Consolidation Type]

## ■ Overview of detaching a BB (2BB to 1BB)

- Release the redundant I/O configuration in the control domain
  - (1) Detach the I/O devices that will be removed from multi-path.
  - (2) Stop using the I/O devices.
- Release the redundant I/O configuration in the root domain
  - (3) Detach the I/O devices that will be removed from multi-path.
  - (4) Stop using the I/O devices.
- Detach BB
  - (5) Run the deleteboard command on XSCF to detach the BB from the PPAR.

## ■ Overview of attaching a BB (1BB to 2BB)

- Attach BB
  - (1) Run the addboard command on XSCF to attach the BB to the PPAR.
- Re-establish the redundant I/O configuration in the root domain
  - (2) Activate the added I/O devices.
  - (3) Attach the I/O devices that are added to multi-path.
- Re-establish the redundant I/O configuration in the control domain
  - (4) Activate the added I/O devices.
  - (5) Attach the I/O devices that are added to multi-path

Please see Building High Availability System on FUJITSU SPARC M12 and Fujitsu M10/SPARC M10 Servers (Maintenance Procedures) for more details.

## 5. Important Notes

## ■ Key configuration for building a BB HA system

- All file systems must be ZFS
- To reduce domain suspend time during PPAR DR, link up all ports on FC cards.
- Do not use physical bindings, specify CID or PA. Otherwise CPU Auto Replacement or PPAR DR will not run.

## ■ XSCF when detaching BB

- When detaching a BB, the XSCF must be a slave or in the standby state(\*). If not, please switch it to a slave or the standby state before executing PPAR DR.

(\*) XSCF resides on each BB. On a multi-BB configuration, one of the XSCFs is the master XSCF, and the others are slave XSCFs. One of the slave XSCFs is the standby XSCF.

See the Fujitsu SPARC M12-2S and Fujitsu M10-4S manuals and product notes for the latest information.

<http://www.fujitsu.com/global/products/computing/servers/unix/sparc/downloads/manuals/>

## ■ Notes when executing PPAR DR

- If a domain is using memory that is on the BB to be detached, the memory addresses may be remapped. If they are, the domain will be suspended.
- Detaching a BB using PPAR DR may not require suspending the domain for the entire time. But, in the following cases, suspending the domain may be required. Suspending domain are as follows.
  - If the kernel memory of each domain exists on the BB to be detached, The domain where the memory exists will be suspended.
  - If the hypervisor's memory exists on the BB to be detached, All domains will be suspended.
- The maximum value of the suspend time is the sum of the following times.
  - Memory movement time
  - Suspend/resume times for on-board devices
  - Suspend/resume times for PCI cards

Please refer to the following manuals for estimation method and details of each time.

- "2.5.2 Considerations When Operating the System for Dynamic Reconfiguration" in the Fujitsu SPARC M12 and Fujitsu M10/SPARC M10 Domain Configuration Guide.
- "Appendix A Cards That Support PCI Hot Plug and Dynamic Reconfiguration" in the Fujitsu SPARC M12 and Fujitsu M10/SPARC M10 Systems PCI Card Installation Guide.
- Suspension causes a delay of the domain's date, so the NTP client should be configured to adjust time automatically.
- Domains that are in the OBP state can cause a PPAR DR failure. Please check whether Oracle Solaris is running or inactive on each domain before executing PPAR DR.
- To reduce the suspension time in PPAR DR, please set the following:
  - The FC port should be linked.

See the Fujitsu SPARC M12-2S and Fujitsu M10-4S manuals and product notes for the latest information.

<http://www.fujitsu.com/global/products/computing/servers/unix/sparc/downloads/manuals/>

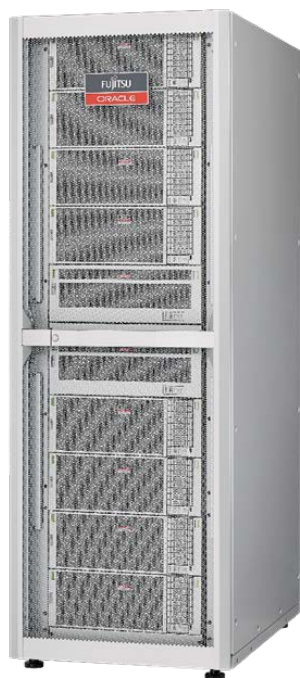
# Reference: Key Fujitsu SPARC M12 and Fujitsu M10 Features for High Availability

## ■ Scalable and Flexible

- Fujitsu SPARC M12-2S server is a modular system that combines Building Blocks for creating a scaled-up server, with up to 384 CPU cores and up to 32 TB of memory.(\*)

(\*) In the case of M10-4S, with up to 1024 CPU cores and up to 64 TB of memory.

- Expand resources without interrupting business and reduce downtime during hardware replacements.

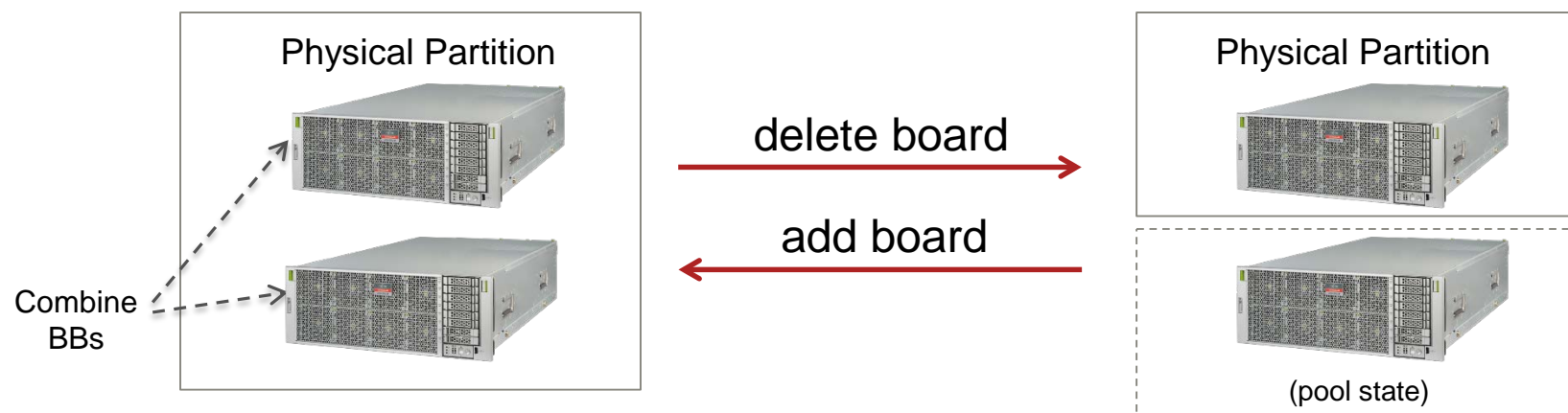


Fujitsu SPARC M12-2S server with multiple Building Blocks.

- ✓ Up to a 16-unit configuration (16 BBs).
- ✓ Gradually adding /removing resources such as CPUs, memory, and PCI slots is a matter of installing additional Building Blocks
- ✓ Each Building Block is dynamically expandable and hot swappable

# Reconfiguring Physical Partitions

- Expanding a system, Physical Partition (PPAR), in BB units
  - A PPAR consists of 2 Fujitsu SPARC M12-2S or 2 Fujitsu M10-4S BBs
  - The system can be reconfigured by adding/removing a BB to/from a PPAR
- Reconfiguring from XSCF
  - The deleteboard command removes a BB.
  - The addboard command adds a BB.



The removed BB goes into the resource pool and can be configured into another PPAR or the same PPAR again.

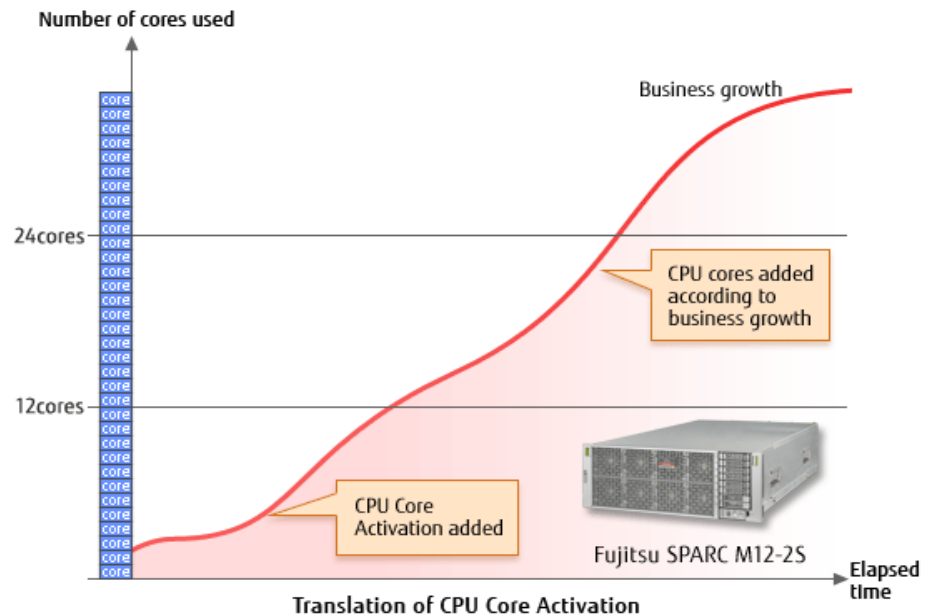


## ■ Reduces initial investment and system upgrade costs

- Fujitsu SPARC M12 and Fujitsu M10 uses CPU core activation. The system can start with a minimum number of activated CPU cores and then be expanded later.
- The minimum number of cores is 2 per Fujitsu SPARC M12-2/M12-2S/Fujitsu M10-1 (4 cores for the Fujitsu M10-4/M10-4S, 1 core for the Fujitsu SPARC M12-1).

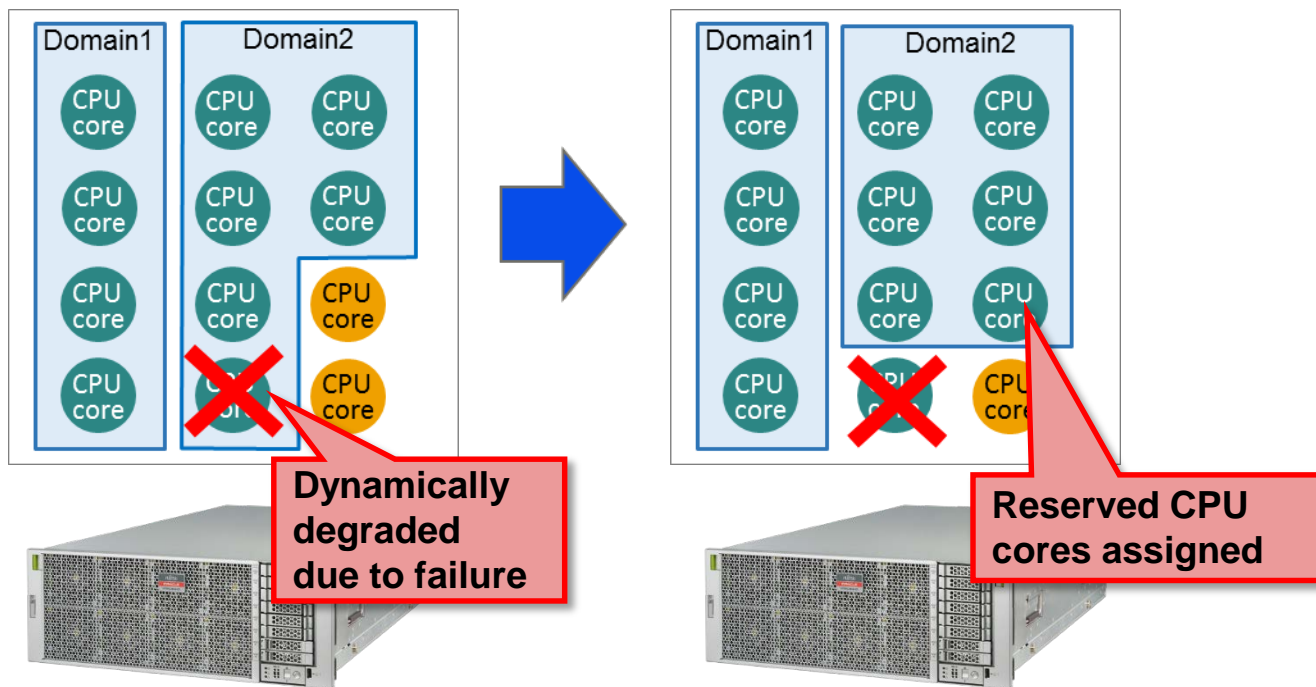
## ■ Upgrade the system anytime as required

- All Fujitsu SPARC M12 models can be expanded in units of one CPU core without system interruption. (Fujitsu M10 models can be expanded in unit of two CPU cores)



## ■ CPU Auto Replacement

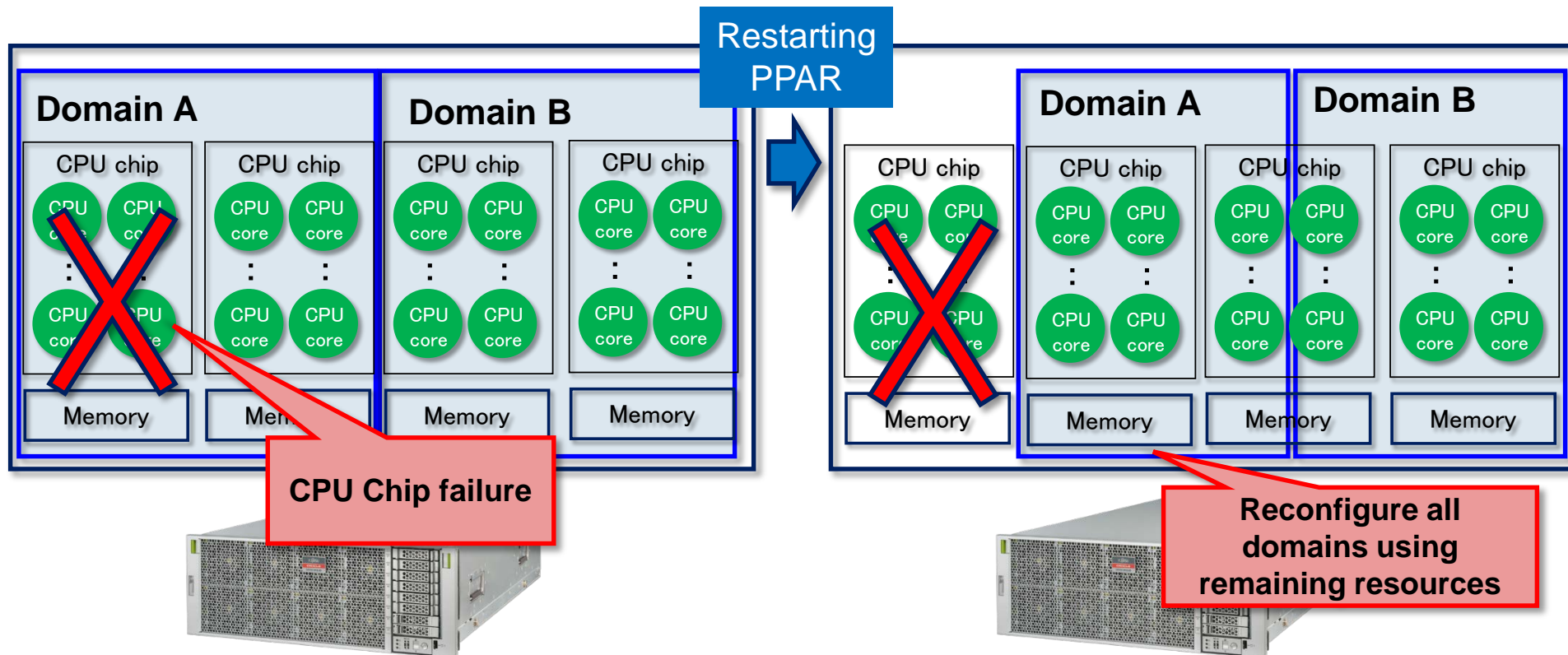
- Un-activated CPU cores are regarded as spare cores. In the event of a core failure, the faulty core is replaced with a spare core, recovering computing power automatically.



# Auto Recovery upon Hardware Failure (2)

## ■ Recovery mode

- If a single CPU chip fails or an entire Building Block fails, Domains (VMs) are automatically reconfigured using the remaining resources and business is resumed.



# PPAR Dynamic Reconfiguration (PPAR DR)

- Reconfiguring PPAR while the PPAR is running
  - Fujitsu SPARC M12 and Fujitsu M10 PPAR DR provides live system reconfiguration.
- Advantages of PPAR Dynamic Reconfiguration

## (1) Expand system resources without interrupting business

- Dynamically add more CPUs, memory, and I/O resources
- Transfer BBs between PPARs according to resource demand

## (2) Reducing downtime during hardware replacements

- Enables BB hot swapping for hardware replacements
- Configuring redundant I/O and using PPAR DR offers minimum downtime for service activities

# Increasing Availability with BBs and PPAR DR

## ■ Minimize downtime upon hardware failure

- Configuring redundant I/O on 2 FUJITSU SPARC M12-2S BBs in the same PPAR or 2 Fujitsu M10-4S BBs in the same PPAR prevents system interruptions due to I/O failures
- If hardware fails, reconfigure domains and recover automatically
- PPAR DR provides Hot replace



**Redundancy**

**Hot replace**

**Auto Recovery**

The system downtime can be minimized by combining Fujitsu SPARC M12-2S and Fujitsu M10-4S Building Blocks and PPAR Dynamic Reconfiguration (PPAR DR).

# Revision History

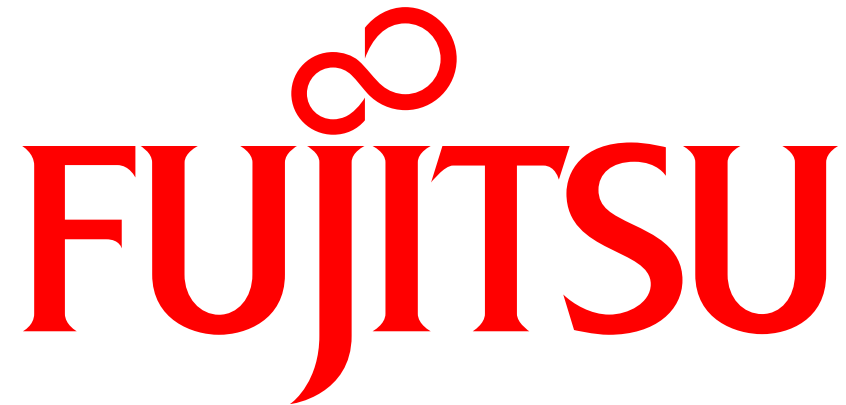
Revision date	Revision	Change
2016.11	1.0	New
2017.4	2.0	Add the Fujitsu SPARC M12-2S Server Update the Hardware requirements
2017.5	2.1	Correction of errors
2017.9	3.0	Add the High Consolidation Type B

## Conditions of use for this document

- About copyrights, trademark rights, and other intellectual property rights  
This content (including text, graphics, sounds, etc.) is protected by copyrights, trademark rights, and other intellectual property rights. This content allows printing and downloading by individuals. But for other purposes (redistributing the content on a website or on any servers), you must receive authorization from our company or the rights holder.
- Disclaimer  
Fujitsu limited and/or its affiliates make no representations or warranties of any kind, whether expressed or implied, regarding this content, which is all provided as is. This content is subject to change or deletion without notice.

## About trademarks

- UNIX is a registered trademark of The Open Group.
- SPARC Enterprise, SPARC64, SPARC64 logo and all SPARC trademarks are trademarks or registered trademarks of SPARC International, Inc. in the United States and other countries, and used under license.
- Oracle and Java are registered trademarks of Oracle and/or its affiliates.
- Other names may be trademarks of their respective owners.



shaping tomorrow with you