Operation Management of Mission-Critical IA Server PRIMEQUEST for TCO Reduction

● Tetsuo Chimoto

(Manuscript received May 20, 2005)

System operation management functions are essential for stable operation of server systems. However, server system operation management is extremely expensive at the moment. The PRIMEQUEST series of mission-critical IA servers uses a server management unit called the Management Board (MMB) mounted in the server cabinet to implement the operation management functions and enables all server management operations to be centrally controlled using the MMB's Web-UI function. In addition, the general-purpose LAN connected in the cabinet eliminates the need to change the LAN cable connections during system installation and configuration changes. These features therefore significantly reduce the cost of server management. This paper describes the various functions and characteristics of operation management that support stable operation of PRIMEQUEST.

1. Introduction

To achieve stable operation of a server system, the following two requirements must be met in addition to high reliability and performance:

1) The operation management functions of the system must be superior.
2) The operation management costs must be low to reduce the Total Cost of Ownership (TCO).

To meet these requirements, the mission-critical IA server PRIMEQUEST uses a server management unit called the Management Board (MMB) to provide the various operation management functions.

This paper introduces the operation management functions and features of PRIMEQUEST.

2. PRIMEQUEST operation management configuration

Operation management of PRIMEQUEST is configured mainly from the MMB and monitor agent software (PSA: PRIMEQUEST Server Agent) installed in each partition (**Figure 1**).

2.1 MMB

The MMB manages all of the hardware in the PRIMEQUEST cabinet.

The MMB manages the various hardware components such as the system boards (SBs), I/O units (IOUs), power supplies, fans, and PCI_Box and the partitions. In addition, the MMB handles various settings such as console redirection, Keyboard/Video/Mouse (KVM) switching, and partition configuration.

The MMB is connected to the hardware components using the two LANs (management LAN and private LAN) described below. Various sensors are also connected using internal buses. The MMB constantly monitors and manages the entire system.

The MMB employs a redundant configuration. If the active MMB fails or an internal network error occurs, the standby MMB is automatically swapped in to continue monitoring of the system. In addition, the MMB can be connected to a remote PC or external operation

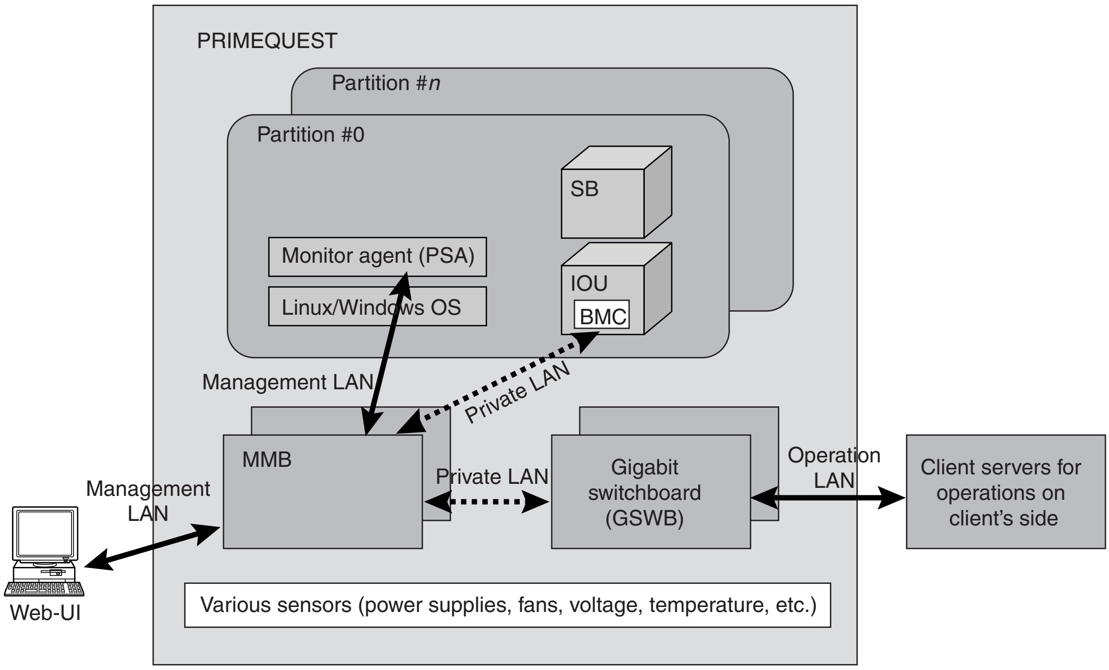

Figure 1
Configuration of PRIMEQUEST server management.

management server via the management LAN. Because the MMB's IP address as viewed from the outside is retained when MMB failover occurs, the MMB's services can continue to be used.

2.2 PSA

Monitor agent software exclusive to PRIMEQUEST called PSA is installed on each partition. The software operates on Linux or Windows to monitor the various hardware configurations and states of the partitions.

The software monitors the CPUs, memories, and chipsets of the SBs and the hard disk drives (HDDs) and PCI slots of the IOUs. If a PCI_Box is connected to an IOU, the software also monitors the PCI card of the connection destination.

3. LAN configuration

In addition to the internal bus connections, connections are made using three LANs within the PRIMEQUEST cabinet. The three LANs are the private LAN, operation LAN, and management LAN. Separate, independent LANs are provided for security and workload distribution purposes.

3.1 Private LAN

The private LAN is used for internal control of the system and is not visible to the OS user. The MMB firmware is linked via the private LAN to the Baseboard Management Controller (BMC) firmware installed in the IOUs that configure the partitions and the Gigabit switchboard (GSWB) firmware that operates on the GSWB.

3.2 Operation LAN

The operation LAN can be freely used for the customer's tasks. The Virtual LAN (VLAN) can be set as required.

The operation LAN is configured from the Gigabit Ethernet (GbE) Controllers installed on the IOUs of each partition and the GSWB connected via the back panel within the cabinet. The operation LAN is connected to the external LAN via the GSWB.

3.3 Management LAN

Unlike other products, PRIMEQUEST has an operation management LAN as standard. This LAN (hereafter called the management LAN) is used to manage the system. The MMB provides external ports of the management LAN for the system administrator, field engineers, and remote maintenance service (REMCS).

Normally, to construct a management LAN for system management, the Network Interface Cards (NICs) and LAN switches must be connected using LAN cables. In PRIMEQUEST, however, NICs and LAN switches for a management LAN are installed as standard. Moreover, these NICs and LAN switches are already wired to LAN cable connections on the back panel, which enables the construction of a cableless environment. The LAN can immediately be used as a management LAN after the server equipment is installed simply by making the minimum required settings, for example, the IP addresses. This not only reduces the number of setup processes, but can also reduce the number of incorrect connections and settings, which helps reduce the TCO.

When a new system board, IOU, or partition is added, only its IP address needs to be set.

Because the GSWB is also connected to the GbE on the IOU via the back panel, as with the management LAN, it is not necessary to connect LAN cables when a partition is added. Operation is enabled simply by setting the VLAN and IP address of the GSWB as required.

4. Operation management functions

Operation management of PRIMEQUEST is realized using the Web-UIs provided by the MMB. The operation management functions of PRIMEQUEST for the MMB are discussed below.

4.1 Web server function

The main feature of the MMB is its internal Web server function. This function provides operation management as a Web-UI. It also enables uniform management of the PSA as the monitor agent of the partitions and enables operation management of the entire PRIMEQUEST server, including the GSWB. As a result, there is no longer any need to prepare an external operation management server to enable the Web-UI. The Web browser functions on a remote PC can be used to perform operation management.

Because the Web server function is not installed on a partition, conflicts with the customer's software environment are reduced. That is, if a customer uses a Web server function such as an Internet Information Server (IIS) or Apache on a partition OS, any operations that are conditional on server management will not be affected. In addition, security patches will of course no longer be required for the Web server software.

Therefore, whatever partition configurations are used, PRIMEQUEST does not need an exclusive operation management server for external use. Instead, operation management of the entire cabinet and each partition can be performed using the Web server function of the MMB and a general-purpose PC.

4.2 High-level application interfaces

For the MMB, the high-level application interfaces provide various APIs required for operation management. An Intelligent Platform Management Interface (IPMI) and Simple Network Management Protocol (SNMP) that conform to world standards are installed. These interfaces can be accessed via the management LAN. In addition, PRIMEQUEST has a Command Line Interface (CLI) for operation management that can be accessed using telnet or the Secure Shell (SSH) via the management LAN.

These interfaces enable linkage to Fujitsu's operation management software Systemwalker and allow the operation management administrator to code scripts in order to automate operations.

4.3 Date and time synchronization

The MMB has an internal Network Time Protocol (NTP). Operating the NTP client on each partition and linking to the NTP server of the MMB enables the date and time of all partitions within the PRIMEQUEST cabinet to be synchronized.

In addition, the MMB itself acts as an NTP client. Moreover, linking to an external NTP server enables the date and time to be easily synchronized between multiple servers.

4.4 Console redirection

The COM port output of each partition can be redirected and accessed from the management LAN via the MMB.

This console redirection enables the console of each partition to be operated from a remote PC without having to operate a telnet server on a partition OS or install an external serial LAN converter.

Moreover, because the console redirection output is buffered within the MMB, the console output can also be accessed from the Web-UI.

4.5 Fault monitoring function

The MMB monitors the various hardware components within the cabinet. If a failure is detected, the MMB stores the failure event in the System Event Log (SEL), which can then be accessed using the Web-UI. In addition, the notification method and destinations can be set in advance to post these failures to the SNMP trap, e-mail, and REMCS center. Moreover, the status of each hardware component can be accessed from the Web-UI. The hardware configuration can be displayed so that the status of each component within the cabinet can be known at a glance.

4.6 Fault-prediction function

To achieve high mainframe-class reliability using the chipsets developed by Fujitsu, PRIMEQUEST monitors the threshold values of correctible errors for the chipsets, CPUs, and memories. If a component whose threshold value has been exceeded within a specific interval is detected, PRIMEQUEST outputs a warning message for the relevant component so a fatal hardware failure does not necessitate stopping the system.

For details about the chipsets, see the paper, “Fujitsu’s Chipset Development for High-Performance, High-Reliability Mission-Critical IA Servers PRIMEQUEST,” presented elsewhere in this special issue.

5. Security considerations

All communications between the MMB and a remote PC or management server can be encrypted (https, SSH, SNMPv3, etc.). In addition, the PSA supports a Web-UI using Web server functions on the MMB. Therefore, Web server functions are not required on partitions, and http ports, which are considered to be security holes, are not required.

For the MMB, access control enables access to the accessible IP addresses and ports to be restricted.

In addition, the LAN switch used for connecting the management LAN installed in the MMB uses a VLAN to further increase security. The LAN switch can stop communications between the partitions or communications from the external ports on the MMB.

These functions therefore ensure maximum security for operation management.

6. High-level operation management software linkage

As described above, the MMB provides IPMI and SNMP high-level application interfaces that conform to world standards. These interfaces enable linkage to Systemwalker and also operation management software of other vendors in order to handle the various operations required by our customers.

Use of the SNMP trap function enables the management server to immediately detect any changes in the hardware or partition status. In addition, the configuration and status within PRIMEQUEST's cabinet can be collected in the Management Information Base (MIB) of the SNMP. The use of these general-purpose functions therefore makes it very easy to link to the operation management software.

Moreover, use of the IPMI further increases the level of settings available to the operation management software.

When Systemwalker is linked to the MMB and PSA of PRIMEQUEST, each state of the MMB, GSWB, and partitions can be displayed on the operation management console of Systemwalker. In addition, the various events that have occurred can be batch-managed and displayed.

7. PRIMECLUSTER linkage function

To achieve the high reliability and high availability required for mission-critical applications, PRIMEQUEST's linkage with clustering software PRIMECLUSTER has been strengthened.

The MMB provides a function for PRIMECLUSTER that monitors the OS status and immediately posts any changes in the OS status and a function that can forcibly stop an OS. These functions enable rapid and reliable failover of PRIMECLUSTER.

8. Conclusion

This paper introduced the configuration, functions, and features related to operation management of the PRIMEQUEST servers.

For the future, we plan to extend the operation management platform based on the MMB, promote linkage with high-level operation management software platforms and autonomous control software for automatic reconfiguration of server resources, and help reduce the TCO of server management.

This research has been partially funded by the Ministry of Economy, Trade and Industry (METI) and the New Energy and Industrial Technology Development Organization (NEDO).

Tetsuo Chimoto received the B.E. degree in Electronics Engineering from Tokyo University of Science, Tokyo, Japan in 1983. He joined Fujitsu Ltd., Kawasaki, Japan in 1983. From 1983 to 1998, he developed network firmware for the Office Server PRIMERGY6000 series. From 1998 to 2002, he developed drivers for the UNIX Server PRIMEPOWER series. Since then, he has been developing server management software for the mission-critical PRIMEQUEST IA server.