



富士通新型高端双核 Itanium PRIMEQUEST 服务器提供最大的高可用性

2006年7月

为 Fujitsu 编制

目录

摘要.....	1
引言.....	2
富士通基于技术的高可用性概览.....	3
高可用性的系统镜像操作.....	3
高可用性系统管理.....	4
热插拔.....	5
集群功能.....	6
更多高可用性功能.....	7
IDEAS 概要.....	7
尾注.....	8

深度研究

本白皮书由Ideas国际 (IDEAS)编制, 详细介绍了有关富士通PRIMEQUEST服务器的高可用性性能。三本姐妹篇白皮书对以下方面提供了类似的详细说明和深度研究: 操作系统、系统架构、以及系统管理。一本概览白皮书提供了有关PRIMEQUEST服务器的功能概要。了解更多详情, 请参见 <http://www.fujitsu.com/global/>。

摘要

富士通最近发布了基于双核 Intel Itanium 2 处理器 (Montecito) 的 PRIMEQUEST 500 系列关键任务企业服务器, 有如下 3 种配置模式: 520 (8 CPU/16 内核), 540 (16 CPU/32 内核)以及 580 (32 CPU/64 内核)。所有 PRIMEQUEST 服务器都运行当前的 Linux 版本(Red Hat Linux 企业版第 4 版 [RHEL4] 和 SUSE Linux 企业服务器第 9 版 [SLES9]) 以及 Windows (Windows Server 2003, Itanium 和 Windows Server 2003 企业版, Itanium 数据中心版)。也可以运行下一个 Linux 版本 (RHEL5 和 SLES10) 以及发布后的 Microsoft Vista。此外, 500 系列服务器可以发挥出 Itanium Montvale 微处理器的最大能力。富士通将这些高端平台称为其运行关键任务的 Intel 架构服务器产品, 定位于那些希望采用高度可靠的企业级服务器, 从业界标准的硬件和软件中获益的客户。

富士通 PRIMEQUEST 服务器与其他 IPF 平台的不同之处在于它对关键任务设计的关注。富士通凭借其大型机和 PRIMEPOWER 的设计经验, 创建出基于 Itanium 的服务器家族, 满足客户在工业标准服务器上运行关键任务负载的需求。

业界公认, 那些具有大型机设计经验的公司如富士通、IBM 和 Unisys 可利用这一经验为设计其他机型带来好处。在这方面富士通尤其突出, 在 PRIMEQUEST 服务器产品中整合了大量类似大型机的高可用性功能。这些功能包括业界独特的系统镜像结构、冗余系统管理、多热插拔功能、集群以及各种其他特点和功能。

系统镜像结构有两种模式: 标准系统镜像和可选系统镜像。每种选项具有不同的功能。系统镜像操作 (取决于选定的选项) 采用富士通大型机技术, 可复制分区资源 (硬件冗余) 并在锁定模式下运行。因此, 任何组件失效均不会导致分区故障。

本档由 Ideas International (IDEAS) 版权所有, 受美国和国际版权法和相关协定保护。未经 IDEAS 书面同意, 不得拷贝、复制、在检索系统中存储、以任何形式传输、张贴于公共或私有的网站或公告板、或授予第三方许可。不得从本白皮书上隐匿或删除版权信息。本白皮书中所指的所有公司和产品的商标和注册商标均受保护。

本档基于认为可靠的信息和来源。本档应“保持原样、不予改变”进行使用。IDEAS 对于内容的数据、主题、质量和时效性不作任何保证和陈述, 并且不承担任何责任。本档中的数据可能有变动。IDEAS 不负责向读者通知数据变更。另外, IDEAS 对于本档所描述的产品、服务和公司的观点可能有所改变。

IDEAS 对于基于本档信息以及读者试图复制性能效果和其他结果所作出的决策不承担责任。本档也不用于预测未来的价格或性能等级。本档不构成本白皮书中所讨论的产品和服务和所讨论的供货商提供的产品和服务的担保。

与业界领先者合作 实现高可用性

为保证 PRIMEQUEST 服务器在高可用性方面达到业界领先的水平，富士通不仅通过基于其长期的大型机优势的自身努力，还与众多功能的主流厂商建立合作关系，如 Intel, Microsoft, Red Hat, Novell, 和 Open Software Development Laboratory (OSDL) 等。经过这些努力，有关各方在进行现有功能改进时会接受和实现富士通的技术建议，其中包括：

- » 富士通提议了 Intel 将承认的 Itanium 增强功能，富士通正与 Intel 合作，为 Itanium 研发类似大型机的机器校验架构，从而增中服务系统的可用性。
- » 富士通正在与 Open Source Development Lab (OSDL) 共同协作，开发的所有高可用性和其他改进功能将成为开放资源。
- » 富士通与 Red Hat 和 Novell 开展合作，他们的发布中将包含富士通的高可用性和其他增强性能。
- » 富士通与 Microsoft 开展合作，以确保 Windows Vista 能够利用 PRIMEQUEST 服务器的高可用性和其他特性。

热插拔功能直接给 PRIMEQUEST 服务器的可用性带来好处。PRIMEQUEST 平台的许多组件在设计上就是可热插拔的。这些组件包括大多数 IO 设备、风扇和风扇托盘、电源、系统管理板、操作面板、键盘/视频/鼠标 (KVM) 单元，和其他系统组件。动态重构将进一步加强该功能。

这份白皮书也讨论了许多其他 PRIMEQUEST 服务器的高可用性功能。包括冗余系统管理板、集群（在两个或更多 PRIMEQUEST 服务器之间或 PRIMEQUEST 服务器内部）等等。此外，富士通正与开源社区和 Microsoft 合作，为 Linux 和 Windows 开发出更多的高可用性功能。

本文详细介绍了 PRIMEQUEST 服务器的高可用性。其他相关文章专注于其他 PRIMEQUEST 特点，如系统架构设计、系统管理功能和操作系统支持等。同时也提供了“PRIMEQUEST 概况”白皮书。

引言

富士通与 Intel、Microsoft、Red Hat 和 Novell 密切合作，基于双核 Intel Itanium 2 处理器 (500 系列) 以 8-CPU/16 内核、16-CPU/32 内核以及 32-CPU/64 内核方式 (分别为 520、540、核 580 型号) 提供高端 PRIMEPOWER 服务器。这些服务器将运行最新和即将发布的 Itanium CPU、以及最新的 Linux 和 Microsoft 操作系统。此外，它们可进行 SMP 或多个分区和域的升级操作。业界领先的富士通高可用性技术将这些功能稳定地联系在一起。

PRIMEQUEST 服务器可在这两种支配市场的操作系统上运行高端应用。根据多数业界分析，Windows 和 Linux 占有 60% 以上的市场。这些服务器与运行 UNIX (SPARC 微处理器上运行的 Solaris 操作系统) 的 PRIMEPOWER 和业界标准的、采用 32-位/64-位 (EM64T) Intel 处理器及 Windows/Linux 的 PRIMERGY 一起，构成了完整的富士通服务器产品的关键部分。

PRIMEQUEST 服务器将提供富士通原创的独特或推荐扩展，延伸到 Itanium、Windows 和 Linux 以进一步改进性能。工具条“与业界领导者展开高可用性合作”概述了富士通在高可用性方面的努力。

建立在富士通长期大型机、矢量处理器 UNIX 设计经验之上的 PRIMEQUEST 服务器可直接在世界市场上与来自 HP、SGI、IBM、NEC 和 Unisys 的机器竞争。主要目标是 HP、IBM 和 Unisys。目标应用包括 OLTP、决定支持/数据仓库、高性能计算、历史移植以及服务器/数据库/应用整合。

富士通与众不同之处：

在众多供应商提供的产品中，富士通以其类似大型机的系统镜像使其服务器分区功能独领风骚，使 Windows/Linux 能够在故障发生时继续工作。

富士通基于技术的高可用性概览

业界公认，那些具有大型机设计经验的公司如富士通、IBM 和 Unisys 可利用这一经验为设计其他机型带来好处。在这方面富士通尤其突出，在 PRIMEQUEST 服务器产品中整合了大量类似大型机的高可用性功能。这些功能包括业界独特的系统镜像结构、冗余系统管理、多热插拔功能、集群以及各种其他特点和功能。

高可用性的系统镜像操作

概览。系统镜像采用富士通大型机技术，可复制（硬件冗余）分区资源（crossbar、存储器）并在锁定步骤下工作。因此，任何组件失效都不会导致分区故障。

16—和 32—CPU 的 PRIMEQUEST（540 和 580 型）服务器均提供系统镜像操作。系统镜像可通过 2 种模式关闭或打开—标准或扩展模式。

关闭情况下所有分区在非系统镜像下操作。要把 PRIMEQUEST 服务器机架改成系统镜像（打开镜像），需要重启机架。在系统镜像打开情况下，每个分区可选择两种模式（标准或扩展）中的任意一种。

需要牢记的富士通系统镜像关键点是当使用系统镜像时，根据选取的模式，crossbar 或存储器故障都不会导致系统分区故障或“跳跃”到一个错误（甚至不会临时中断恢复）中。其他硬件厂商—包括 HP、IBM 和 Unisys，在其 Windows/Linux 产品中都不具备这种类似大型机的功能。这是 PRIMEQUEST 服务器与其竞争产品的区别之处。

系统镜像打开时，每个 PRIMEQUEST 服务器分区默认的镜像模式都是标准镜像。扩展镜像模式需要另外再打开。如上所述，系统镜像打开时，所有分区可分别设置为标准或扩展模式。

模式细节。标准系统镜像只为双 PRIMEQUEST 服务器地址 crossbar 单元提供硬件冗余。PRIMEQUEST 服务器的富士通芯片组中也为地址接口电路提供冗余 (PRIMEQUEST 500 系列的新功能)。

在扩展系统镜像中，地址 crossbar 单元（同步操作）、四向数据 crossbar 单元（同步操作）、所有主存（重复的存储器读/写访问）以及 PRIMEQUEST 服务器富士通芯片组均提供硬件冗余配置。标准系统镜像功能（如地址 crossbar）包含在扩展系统镜像中。

通过 PRIMEQUEST 系统镜像，用户可以根据其应用的需要权衡性能和可用性。

表 1 (如下) 提供富士通硬件冗余的系统镜像高层概览。

	系统镜像		
	非镜像模式	标准	可选
本地数据 Crossbar 及存储器	非镜像	非镜像	镜像
数据 Crossbar	非镜像	非镜像	镜像
地址 Crossbar	非镜像	镜像	镜像

表 2 (如下) 显示标准系统镜像和可选系统镜像对 PRIMEQUEST 服务器性能的影响效果。由于性能和系统镜像选择相关，因此当应用功能运行时，正确测量性能和所有 PRIMEQUEST 分区的可用性之间的权衡十分重要。

打开扩展系统镜像时，每个分区有效的内存容量减少一半（扩展镜像下，内存插槽分成 2 个内存镜像组，这些组之间进行镜像操作。逻辑内存容量变为物理容量的一半）。类似的，PRIMEQUEST 分区的系统板和 IO 板的 crossbar 带宽也将减少，如图 2 所示。

系统镜像带来的任何分区性能降低都可由下面的情况来弥补，即故障发生时系统会继续运行，很少或不会给业务单元经理和端用户带来损失。通过系统镜像操作保证最大的高可用性时，如果需要更大的应用容量，可以添加更多的资源扩大分区。

表 2. 系统镜像硬件冗余覆盖率及 PRIMEQUEST 500 系列服务器的影响

项目	系统镜像		
	关闭	打开	
		标准镜像	可选镜像
全局地址总线	单个	双倍	
全局数据总线	单个	双倍	
内存	单个	双倍	
有效 MEM 容量	100%	50%	
有效性能	100%	>99%	>95%
Xbar B.W.(SB)	17.06 GB/秒	8.53 GB/秒	
Xbar B.W.(IOU)	4.26 G B/秒	2.13 GB/秒	

高可用性系统管理

高级系统管理对高级可用性十分关键，如果无法根据需要获知系统状态，或无法对系统状态进行更改（如修复），则该系统在高可用性方面就不是值得信赖的系统。了解实际系统操作情况后，富士通 PRIMEQUEST 服务器提供了集成的系统管理板。在 540 和 580 型号的冗余配置中得到支持，不仅针对单点故障。此外，许多系统管理板的组件都是冗余备份的，以获得更高可用性。

富士通 PRIMEQUEST 服务器通过系统管理板提供多个分区的整合管理。该功能进一步确保系统管理员可优化 PRIMEQUEST 服务器的正常使用时间。

作为整合管理的一部分，在需要维护时可动态更新分区并进行并发维护。该功能的完全实现要等 2006 年后期或 2007 年发布新版本的 Red Hat (RHEL5 和 6) 和 Novell SUSE (SLES11)。动态重构可用也要等 Microsoft 发布 Vista 操作系统。关于该主题更详细的信息请参看有关操作系统的相应白皮书。文档后面也对动态重构作了进一步探讨。

有关系统管理板的更多信息可在系统管理系列的相关白皮书中找到。然而，重要的是系统管理板也提供改进 PRIMEQUEST 高可用性的许多其他功能。系统管理板：

- » 推动自主故障恢复，使得停机时间达到最小或避免停机。
- » 使用一百多个 GUI 屏幕进行系统监控，通过 web 浏览器连接的系统管理板控制台进行监控。控制台提供所有 PRIMEQUEST 服务器分区组件的状态数据和其他系统级信息数据（直到机架组件温度级别）。这样细致的监控确保停机时间达到最小或避免停机。
- » 带来控制 PRIMEQUEST 的灵活 IO 开关。通过灵活的 IO，IO 资源可根据需要独立灵活地连接到 CPU 和相应的内存资源。这种功能的优点是可快速从系统板故障中恢复，进一步增加系统高可用性。

热插拔

热插拔功能对服务器可用性十分关键，许多 PRIMEQUEST 组件在设计上都是可以热插拔的。这些组件包括 I/O 设备、风扇和风扇托盘、电源、系统管理板、操作面板、KVM 单元，以及其他系统组件。

关于哪些从根本上可以热插拔的详细分类取决于 PRIMEQUEST 500 系列的型号、取决于 PRIMEQUEST 客户选择了何种冗余选项以及使用了何种操作系统。表 3 是 500 系列哪些 PRIMEQUEST 组件是冗余和/或热插拔的高层视图。

表 3. 备份和热插拔的 PRIMEQUEST 组件

	Model 520	Model 540	Model 580
冗余	SB, PSU, FAN	SB, GSWB, MMB, PSU, FAN	SB, GSWB, MMB, PSU, FAN
热插拔	SB, IO, IOX, PSU, FAN, PCI Card, HD	SB, IO, GSWB, MMB, PSU, FAN, PCI Card, HD	SB, IO, GSWB, MMB, PSU, FAN, PCI Card, HD

SB=系统板; PSU=电源单元; GSWB=千兆交换板; MMB=管理板; IO=输入/输出单元; IOX=高速输入/输出单元; HD=硬盘驱动器

除非不在工作分区中，否则许多主要组件基本上就不支持热插拔。如果支持动态重构，则运行分区中的主要组件可进行热插拔。当然，运行分区外的系统板和 IO 单元在替换时不会影响运行分区。单独的系统板 CPU 和 DIMMS 只有在把系统板从 PRIMEQUEST 服务器机架中拆除后才能更换。

动态重构改进。动态重构功能出现后，上述热插拔功能就更容易实现。比如使用动态重构后，系统和 IO 单元可以热交换，正如某些 crossbar（如运行分区中的数据 crossbar）那样。有些问题组件可能需要重启分区。此外，动态重构可能需要冗余路径（如 IO）。有许多其他变量要考虑。下面的列表概述了当前富士通 PRIMEQUEST 在这些领域的改进计划(最佳情况)。

- » PCI-X 设备热插拔
- » PCI-EX 设备热插拔
- » 热添加系统板
- » 系统板热删除
- » I/O 单元热插拔

哪些组件可以热交换、或需要动态重构进行热交换、以及是否需要重启等更详细的讨论都超出了这份白皮书的范围。这些细节可在富士通的 PRIMEQUEST 文档中找到。新版本 Windows 和 Linux 出现后，这些信息会快速更新。最新细节可通过富士通得到。

最后，谈到动态重构时，要记住在需要时，动态重构可使 PRIMEQUEST 系统管理员和用户体验并发维护的好处。这意味着系统资源可根据需要在系统中移进移出，不会对应用带来影响。

集群功能

PRIMEQUEST 服务器可集群构建，满足关键任务环境的高可用性要求。这种集群可通过两个（或更多）分开的 PRIMEQUEST 单元实现，或者单一 PRIMEQUEST 服务器可配置成内部集群结构。在两种情况下现有的 PRIMEQUEST LAN 连接均可提供必须的时钟信号、系统管理板信息、用户信

PRIMEQUEST 500 系列服务器为双核 Intel Itanium 2 微处理器上运行的 Windows 和 Linux 系统提供了一个功能强大、可用性高的物超所值的选项方案。

息等等。

需要强调的是 PRIMEQUEST 镜像并不等同于集群，也不能替代集群。二者之间有许多不同，结构也满足不同的要求，此处不详细介绍。集群可适应应用故障并进行切换，需要故障切换时间。而镜像意识不到应用服务的中断，并且没有故障切换时间。

富士通 PRIMECLUSTER 产品为 PRIMEQUEST 500 系列服务器提供必要的集群软件。PRIMECLUSTER 是工作在 Solaris 上经过长期检验的产品，也移植到了 Linux 上。当采用 Windows 作为操作系统时使用 Microsoft 的 MCS 软件。

更多高可用性功能

虽然此处没有详细讨论，但 PRIMEQUEST 服务器还有其他高可用性功能，可确保停机时间最小。这些功能包括（还有许多其他的）：

- » 包括冗余控制电路的富士通芯片组高可用性功能
- » 使用服务器内部的虚拟 LAN(VLAN)，避免系统和 IO 板以及外部接口间缠绕在一起的线缆
- » 在可能的地方（包括 ASIC）广泛使用差错检查和纠正（ECC）以及偶校验
- » 协议检查，保证正确的内部通信
- » CPU 和存储器故障预测，拆除发生故障的组件
- » PCI 总线故障检测
- » 内建的双核 Intel Itanium 2 处理器 (Montecito)的高可用性功能

IDEAS 概要

富士通的 PRIMEQUEST 500 系列高端、关键任务服务器产品线基于双核 Intel Itanium 2 处理器 (Montecito)构建，专门支持 Windows 和两个主要的 Linux 版本，PRIMEQUEST 服务器支持多种垂直产业应用，客户在关键任务平台上可拥有工业标准的硬件（Intel）和操作系统（Windows 和 Linux）。

Windows 和 Linux 无需修改就可以在 PRIMEQUEST 上使用。这些操作系统已经让 PRIMEQUEST 可提供各种高可用性功能。然而，富士通正与 Microsoft 和开源社区紧密合作，在 Windows 和 Linux 上添加新功能，使其更加适合 PRIMEQUEST 服务器（和其他计算机系统）。尤其在高可用性方面，富士通正与 Microsoft 和 Linux 社区合作开发 dump 软件、动态重构功能，改进高可用性的集群功能，机器检查结构以及 RAS 功能（如热插拔）。

富士通也正在开发自己独特的 PRIMEQUEST 管理功能和高可用性的芯片组。

除了本白皮书提到的高可用性功能，富士通也在其他方面对 PRIMEQUEST 500 系列进行改进。此处由于篇幅限制不详细讨论，这些改进包括：

- » 富士通现场测试数据表明，与早期 PRIMEQUEST 型号相比，MTTR 减少了 90%
- » 由于没采用电缆，而是采用中型板的无线电缆设计，增加了可靠性，降低了成本 (COO)
- » 非单点故障结构的进一步开发
- » 采用浮动系统板代替故障板 (基于灵活的 I/O 功能) 的能力
- » 降低整体系统功耗，减小热量带来的器件损坏
- » 类似大型机的校验结构，满足 PRIMEQUEST 中的双核 Intel Itanium 2 处理器

上述功能与本文中提到的所有高可用性功能相结合后，PRIMEQUEST 服务器将占领各种应用市场。这些应用存在于高端、关键任务服务器 Linux 或 Windows 环境（或二者兼有），该环境中高可用性是主要因素。

由于架构原因及其所带来的高可用性，富士通预期 PRIMEQUEST 服务器将能够提供与其他厂商的基于 Itanium 的产品相比更高的 COO 和投资回报 (ROI) 率。该投资效率与 PRIMEQUEST 服务器的高可用性功能以及其他优势一起 (在本系列白皮书中已进行重点讨论)，保证全球客户在了解 PRIMEQUEST 的独特位置和价值比例后一定会对它产生兴趣。

尾注

若要更深入了解本节材料，请参考相关白皮书，“PRIMEQUEST 系统架构,” Ideas 国际,2006 年 7 月。

参看相关白皮书，“新的富士通双核 Itanium 2 PRIMEQUEST 服务器为 Linux 核 Windows 应用提供关键任务的主机系统,” Ideas 国际,2006 年 7 月。

参看相关白皮书，“新的富士通双核 Itanium 2 PRIMEQUEST 服务器继续提供业界领先的系统管理,” Ideas 国际,2006 年 7 月。

4 参考 “Montecito 错误防护与降低”，来源于 Intel。

Americas

Ideas International, Inc.
800 Westchester Avenue
Suite S620

Rye Brook, NY 10573-1330
USA
Tel + 1 914 937 4302
Fax +1 914 937 2485

Asia/Pacific and Worldwide Headquarters

Ideas International Limited Level 3

20 George Street
Hornsby, NSW, 2077
Australia

Tel +61 2 9472 7777
Fax +61 2 9472 7788

Europe, Middle East, Africa

Ideas International Europe 1
Deanes Close

Steventon
Oxon OX13 6SZ
United Kingdom

Tel +44 (0) 1235 437 850
Fax +44 (0) 1235 437 851

www.ideasinternational.com

