

PRIMEQUEST系统架构

2006年七月



为Fujitsu编制

目录

摘要	1
引言	2
客户需要高端业界标准关键任务服务器 ...	2
PRIMEQUEST 系列	2
PRIMEQUEST 500 系列	3
以关键任务设计致胜	5
IPF Processors的优点	5
富士通芯片中的差错校验与修正	5
镜像总线、内存	6
线缆冗余提升可靠性	7
配置灵活性	8
分区粒度	8
并发维护性能	9
IDEAS概要	10
尾注	11

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 服务器的系统架构。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: 高可用性、操作系统、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

随着Linux和Windows操作环境以及采用Intel处理器的服务器吸引了越来越多的客户, 富士通于2005年发布了采用Intel的Itanium (安腾) 2处理器的PRIMEQUEST 400系列。PRIMEQUEST产品线通过采用双核Intel Itanium 2处理器 (以前的代号为“Montecito”) 进行了更新。双核处理器和富士通独创的新特征增强了新的PRIMEQUEST 500系列型号的可扩展性。PRIMEQUEST是为满足企业客户的需要设计的, 为这些客户提供了一系列高度可靠的服务器, 这些服务器可以从业界标准软硬件获得经济方面的收益。

根据定义, 术语“业界标准”意味着还有其他系统也采用Itanium芯片运行Linux或Windows。富士通的服务器与其他基于IPF的平台之间的区别是, 特别注重关键任务设计。富士通利用大型机和PRIMEPOWER设计经验, 开发出了基于Itanium的系列, 满足了客户在使用流行的操作环境的业界标准服务器上运行关键任务负荷的需求。

除了广泛应用错误校正码 (ECC) 之外, PRIMEQUEST服务器还为镜像关键组件提供了多种选择以确保高可用性。实现系统互连的地址和数据crossbar交换机采用冗余设计, 以便它们能配置为携带双重信号。通过镜像的冗余路径, 即使crossbar交换机的一部分出现故障, 系统也能继续运转。内存模块也能采用冗余配置。

由于可以扩展到32个处理器 (具有64个内核), 所以PRIMEQUEST可以处理大量应用负载。对于不需要将所有的32个处理器芯片都专用于单个应用的用户, PRIMEQUEST服务器提供了灵活的分区。分区利用下层的模块构件方式, 由具有处理器、内存和I/O单元的系统板组成, 提供PCI插槽和G比特以太网。最多可配置16个独立的分区, 每个分区都运行自身的操作系统实例: 各种版本的Linux或Windows, 每种操作系统都具有定制的软件栈。每个独立的分区都与其他分区的硬件或软件故障之间实施电隔离。

富士通实现的一个重要特征是, PRIMEQUEST分区允许独立于系统板分配I/O单元。PRIMEQUEST灵活的I/O (Flexible I/O) 允许将I/O资源与任何系统板关联起来, 使得可以通过适当地整合处理能力和I/O连通性来定义分区。

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去除版权信息。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。

总而言之，富士通利用设计企业级服务器的长期经验开发了它的关键任务IPF服务器。即使是不熟悉以前的富士通高端系统的客户也会认识到，PRIMEQUEST平台提供了独特的特征，这些特征会吸引那些需要将业界标准服务器用于关键任务应用的客户。本文的正文部分更详细地描述了PRIMEQUEST服务器设计的一些关键架构特征。其他姐妹篇白皮书着重描述其他的PRIMEQUEST特征，如操作系统支持、高可用性特征和系统管理功能。“PRIMEQUEST 概要”白皮书也可供参考。

引言

客户需要高端业界标准关键任务服务器

以开放的、经过彻底检查的Linux操作环境作为关键任务计算的一种安全的、鲁棒的解决方案逐渐地吸引了许多客户。其他企业客户则很重视适用于被广泛采用的Windows环境的大量应用软件。这些Linux和Windows用户都同样会被业界标准处理器驱动的服务器吸引，因为这是避免受限于单个厂商的一种方式，这要感谢各个厂商提供的多种平台选择。

但是，通过使用业界标准处理器向客户提供多种平台的选择，给厂商带来了一种挑战——如何在遵循这些业界标准的同时使得他们的系统卓尔不群。对富士通而言，这种区别来自于关键任务服务器设计的长期经验。本文以及这个系列中的其他姐妹篇白皮书强调了PRIMEQUEST服务器区别于其他支持Windows和Linux的业界标准平台的特征。

PRIMEQUEST系列

PRIMEQUEST系统基本上是由模块化构件构成的可扩展SMP（对称多处理器），模块化构件由系统板、I/O单元、crossbar和系统管理模块（System Management Module）组成。每个系统板（system board, SB）都包含四个处理器芯片和32条插内存的DIMM插槽。系统板连接到crossbar交换机上，交换机允许任何处理器访问所有的内存，提供了将处理器连接到I/O子系统的通道。图1（左）描述了PRIMEQUEST架构的高层方框图。

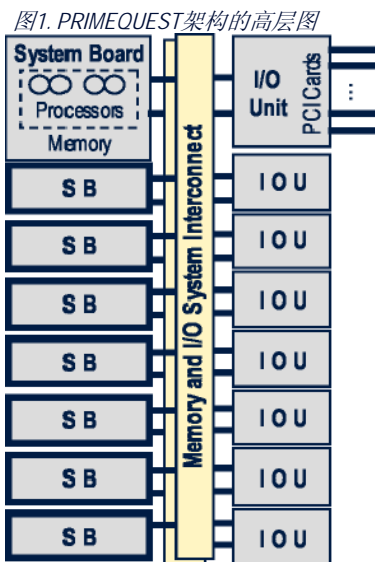


图1在某些方面表现得类似于其他模块化可扩展SMP的高层方框图，但两者之间存在一些重要区别。例如，互连的crossbar是整个系统性能的关键性决定因素。富士通利用它的超级计算机和大型机经验设计了crossbar交换机，能以低于竞争系统的延迟提供更高的带宽。也就是说，与其他可能使用相同的处理器芯片的实施方案相比，PRIMEQUEST提供了更高的系统性能。

值得注意的另一点是系统板和I/O单元可以独立地分配给不同的分区。与其他厂商不同，富士通的方式不是预先确定好哪些I/O单元分配给特定的系统板，而是允许分区灵活地包含不同数量的I/O单元。后面的章节进一步说明了PRIMEQUEST分区特征。后面的章节还详细说明了富士通设计的其他独特的性能，如镜像。

为简便起见，图1没有显示专用的系统管理板（也称为MMB），它们监控服务器的状态，提供配置控制，包括分区。由于在一个嵌入式PowerPC 440处理器上运行实时Linux内核，所以系统管理板中包含的服务器管理功能不需要系统控制台或者专用控制台。图1还省略了可选的G比特以太网交换机板，它提供了一个集成的交换机，用于提供与I/O单元的G比特以太网连接。

图1的高层图示描述了PRIMEQUEST服务器的公共特征，包括已有的400系列和新型的500系列。下面几个段落概述了PRIMEQUEST机型与将来计划的机型之间的区别。

PRIMEQUEST 500系列

强大的处理器是高端系统的基础，富士通在PRIMEQUEST中采用Intel的尖端Itanium 2处理器。最初的型号发布于2004年，它们采用了单核Itanium 2处理器（称为“Madison 9M”，芯片上最多可以容纳9M高速缓存）。这些400系列型号可以配置成支持8、16或32个Itanium 2处理器芯片。随着Intel的双核Itanium 2处理器芯片（以前的代号称为“Montecito”）的发布，富士通现在推出了PRIMEQUEST 500系列，它们也可以配置成8、16或32个Itanium 2处理器芯片，但与以前的型号相比，容纳的处理器内核增加了一倍。表1总结了PRIMEQUEST系列。

表1. 400系列和500系列型号

PRIMEQUEST 型号	处理器芯片 处理器类型	最大内存
580	64 cores Dual-Core Intel Itanium 2 processor	2 TB 1.6 GHz
540	32 cores Dual-Core Intel Itanium 2 processor	1 TB 1.6 GHz
520	16 cores Dual-Core Intel Itanium 2 processor	256 GB 1.6 GHz
480	32 cores single-core Itanium 2	1 TB 1.6 GHz
440	16 cores single-core Itanium 2	512 GB 1.6 GHz
420	8 cores single-core Itanium 2	256 GB 1.6 GHz

更多的内核可以处理更多的吞吐量；但是，内核数量不是影响系统性能的唯一因素。表2比较了最高端型号480和580之间最重要的区别和相似之处。

	480型	580型
处理器	32 单核 Itanium 2处理器	32 超线程双核Itanium 2处理器
时钟/ 缓存	1.6 GHz / 9 MB, 1.5 GHz / 4 MB	1.6 GHz / 24 MB, 1.42 GHz / 12 MB
前端总线	400 MHz	533 MHz
每块系统板上的处理器芯片	4 个单核芯片	4 个8核芯片
内存	1 TB DDR2 (4 GB DIMMs)	2 TB DDR2 (8 GB DIMMs)
Crossbar带宽 (峰值)	103 GB/sec.	137 GB/sec.
内部硬盘	4.7 TB (32 drives @ 147 GB)	4.7 TB (32 drives @ 147 GB)
可选I/O 组件	8 each with 12 PCI-X slots	8 each with 12 PCI-X slots
PCI 插槽	128 (32 internal, 96 in I/O boxes)	128 (32 internal, 96 in I/O boxes)
分区	8 4个接口粒度	16 2个接口粒度

表2. PRIMEQUEST 480与580型号之间的比较

注意，双核处理器芯片包含大量芯片高速缓存（on-chip cache），这有助于提升性能。双核Intel Itanium 2处理器“Montecito”从单核“Madison 9M”采用的130纳米工艺提升为90纳米加工技术。更小的晶体管使得Montecito芯片可以容纳17.2亿个晶体管，而Madison 9M芯片中只有5.92亿个晶体管。因此，Intel可以在每个芯片上放置两个完整的Itanium 2处理器内核，而且可以将每个内核的L3缓存增加到12MB。此外，双核Itanium提供了称为超线程（Hyper-Threading）的多线程方式，允许每个内核并行地执行两个不同代码线程的指令。

增加的L3缓存应该有助于缓解两个内核对公共前端总线的竞争。此外，当一个线程在等候内存中的数据的时候，多线程也应该会使得执行单元处于忙状态。多线程的优点很大程度上取决于运行的代码，但能够将性能提升10%至20%。双核、大缓存和超线程的组合大致上使得双核“Montecito”Itanium 2芯片提供的吞吐量最多可以达到单核“Madison 9M”芯片的2.5倍，其应用范围包括通常在大型SMP服务器上处理的事物处理工作。此外，双核Intel Itanium 2芯片的吞吐量性能的提升不需要额外的电能。实际上，Intel将“Montecito”双核芯片设计得比单核“Madison 9M”消耗更少的电能（因此冷却的需求更低）。

表2也表明，前端总线在双核芯片上运行得更快，为内存、I/O、和PRIMEQUEST系统互连提供了更大的带宽。富士通重新设计了它的支持芯片组，以利用更快速的传输速率，从而提升了表2中所示的crossbar带宽。

PRIMEQUEST的高性能crossbar系统互连是它区别于其它基于IPF服务器的优点之一。PRIMEQUEST的crossbar将Itanium 2处理器组合在一起创建了一种共享内存的多处理服务器。富士通工程师利用他们的大型机和超级计算机技术，设计出了高带宽、低延时的crossbar，提供了高达每秒137G（峰值）的总线带宽，超过了竞争对手。

除了“Montecito”之外，Intel还计划开发一种更强大的双核芯片（开发代号为“Montvale”），目前期望在2007年下半年发布。尽管关于Intel未来芯片的具体细节还不清楚，但富士通已经将PRIMEQUEST系统设计得可以采用未来的多核Intel Itanium 2处理器，满足客户对投资保护的需求。

以关键任务设计致胜

IPF处理器的优点

Intel设计了Itanium处理器与最好的RISC芯片进行激烈竞争，其竞争范围不仅包括性能还包括内置的故障恢复（resilience to failures）能力。目前，Itanium的大容量L3缓存是通过ECC进行保护的，ECC能修复单个比特引起的故障。Montecito将增加一种技术，Intel称之为Pellston技术，它允许禁用缓存中出现故障的部分。Montecito还将包括SpeedStep技术，该项技术首先用在移动处理器中，它能在处理器的安静期间（lull period）降低时钟频率以减少功耗。除了节约电力成本之外，SpeedSte还有助于双核Montecito保持更低的温度，从而可以提升电路的可靠性。总的来说，Intel指出，它增加了许多差错检查特征，作为增强Itanium的MCA（Machine Check Architecture，机器检查体系）的一部分措施。尽管增强MCA的具体细节仍然没有完全公布，但Intel报告说，它在某个国家实验室测试软件错误恢复期间，用伽玛射线和 α 粒子冲击Montecito。该项加速压力测试指出，单个芯片的平均失效时间长达数百年。

Montecito中的17.2亿个晶体管还允许Intel添加功能，如Intel的虚拟化技术（Virtualization Technology）。Intel虚拟化技术的正式代号为Vanderpool，它允许系统以同一个处理器在不同的虚拟分区中更好地运行多个操作系统。请注意，Itanium芯片中的Intel虚拟化技术是补充（而不是取代）PRIMEQUEST服务器提供的物理分区（稍后讨论）。

富士通芯片中的差错校验与修正

除了整合在Itanium处理器内的ECC电路之外，富士通还在它的工程师设计的PRIMEQUEST芯片组中包含了ECC保护。具有ECC保护的两个主要范围是crossbar交换机中的系统互连部分和内存接口部分。

系统互连部件实际上由六个独立的crossbar交换机组成：四个处理数据，两个处理地址。这些系统互连都是通过ECC电路进行保护的，可以实时地完全纠正单比特差错，不会影响系统运转。多比特差错可以被检测到，但会导致一次机器检查，因为不能自动纠正这些错误。如果系统管理板处理器确定一根互连总线已永久损坏，就可以使那根总线无效。

内存不仅会有单比特差错，还有会使整个芯片无效的故障。PRIMEQUEST内存通过一种高级ECC编码（称为S4EC-D4ED²）对此进行保护。这种16比特的ECC编码通常被描述成“chip kill”技术，它不仅能纠正单个差错，还能从任何出故障的内存芯片重构所有的四个比特。

总的说来，富士通采用先进的ECC来保护系统中最容易出现错误的部分。

镜像总线、内存

尽管富士通在已经高度可靠的Intel处理器之上另外整合了传统的ECC，但仍然有关键任务客户环境需要更高等级的可用性。对于这种情形，富士通提供了一系列镜像选择，这是其他任何高端服务器所无法匹敌的。特别地，PRIMEQUEST系统为镜像总线或内存提供了多种选择，确保在出现故障时也能继续运转。除了以非镜像配置运行之外，还为PRIMEQUEST服务器提供了两种模式——标准和可选。

当配置为镜像时，资源都是成对的，并行执行相同的功能。如果资源对中的一半出现问题，另一半资源仍然会提供有效的结果，允许系统继续运转而无需中断。如表2所示，标准系统镜像先镜像全局地址crossbar，这些crossbar作为所有的系统板和I/O单元之间的地址互连部件。可选系统镜像将四个全局数据crossbar处理成两份相同的镜像并且配置内存，以便每个存储单元都被复制到两个独立的内存模块中。

表3: PRIMEQUEST服务器提供的系统镜像

	标准系统镜像	可选系统镜像
地址总线	镜像	镜像
数据总线	独立	镜像
SB带宽	12.8 GB/sec. (400 Series) 17.06 GB/sec. (500 Series)	6.4 GB/sec. (400 Series) 8.53 GB/sec. (500 Series)
IOU带宽	3.2 GB/sec. (400 Series) 4.26 GB/sec. (500 Series)	1.6 GB/sec. (400 Series) 2.13 GB/sec. (500 Series)
内存	Independent	Mirrored
内存容量	100%	50%
性能	>99%	>95%

全局地址crossbar将来自于内存或所有系统板I/O的所需数据的地址进行广播。未检测到的错误可能导致数据被写到错误的地方，随后，这可能会发展成严重的系统故障或者成为受损数据。PRIMEQUEST地址crossbar实际上由两根完整的地址总线组成。每个系统板都有单独的芯片连接每根地址总线。在非镜像操作中，两根地址总线都被使用。但两个总线上的总体预期负载非常低，可以只使用一根总线，另一根总线作为镜像拷贝；根据富士通的预测，总体性能的下落不足1%。因此，标准系统镜像（Standard System Mirror）是PRIMEQUEST服务器上的默认镜像方式。

扩展镜像（Extended Mirror）通过复制全局数据crossbar和主存储器进一步采取了一些措施。全局数据crossbar由四个单独的数据通道组成。每个通道都由单比特的差错修正ECC进行保护，系统和I/O板上都由专用的接口芯片。但对于需要特别的保护程度的客户而言，四个数据总线可以配置为两个镜像对。因为当数据总线被镜像时，系统板接口电路的运转状态也是镜像配置，所以保护了系统，避免了接口电路和总线本身的故障。

更重要的是，扩展镜像复制了主存储器中存储的数据。正如前面所提及的，内存使用了ECC编码技术，它不仅能纠正单个比特错误，还能从有问题的内存芯片中重构所有的四个比特。但是，如果出现了额外的多个错误，就不能恢复内存单元的内容。内存镜像将数据复制在两个单独的内存芯片组中。可能导致一个内存单元无效的多比特故障不太可能同时发生在镜像单元中。此外，镜像接口电路还避免了电路故障引起的系统故障。因此，镜像内存提供了强大的数据完整性，这是极度苛刻的关键任务应用所需要的。当然，镜像内存将总的内存容量减少了一半。由于可以在PRIMEQUEST服务器中配置大量内存，所以总的内存容量的减少应该没什么关系。尽管如此，将数据存储在本副本内存单元确实会使内存成本加倍。因此，可选系统镜像要安装容量加倍的内存，这会带来财务成本。就性能损失而言，富士通预计，镜像数据总线的确实会降低可用的crossbar带宽，但导致的性能损失不到5%。

线缆冗余提升可靠性

扩展为许多个处理器的服务器，如PRIMEQUEST，并不总是在一个操作系统的单个实例的控制下以单一的共享内存配置运行。大型服务器通常会被分区成更小的子集以充当服务器整合平台。每个子集都运行它自身的软件，如同一个独立的服务器一样。这种独立性要求每个分区都独立地连接数据LAN，通常还要独立地连接一个与管理系统相连的独立的LAN（以及备份路径）。此外，每个分区通常还需要键盘/视频/鼠标（KVM）连接。

对于高端PRIMEQUEST系统中可用的16个分区，富士通推测，所有的连接所需要的线缆将超过100根。此外，外置的以太网连接需要占用PCI插槽。外部的LAN和KVM交换机增加了额外的成本。如此多的物理线缆的可靠性很可能会出现问

PRIMEQUEST解决方案为集成的、冗余的用于G比特数据LAN的交换机、系统管理板交换机和KVM提供了内置的预连线连接。因为可以灵活地重新配置分区，所以不需要重新接线缆。与物理线缆和外置交换机的硬件成本、安装成本和维护成本相比，这种集成化方式节约了总成本。此外，内部连接也比外部线缆更为可靠。

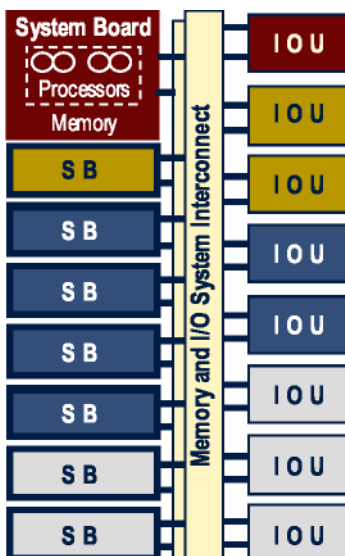
配置灵活性

没有多少单独的应用，尤其是那些运行在Linux或Windows中的应用，会需要一个与64内核PPRIMEQUEST 580 SMP一样大的系统。与此同时，客户将大型系统当作工具，用于将多个服务器的工作负载合并到一个易于管理的中央服务器中。有些应用软件不能融洽地共处，因为它们可能具有不相容的软件栈；例如，它们可能依赖于不同版本的中间件软件。在这种情形下，大型的整体服务器就需要进行分区，划分为更小的操作系统映像。分区提供的故障隔离也非常有价值，它确保了处于开发或测试阶段的软件中的故障不会影响实际负载。富士通根据经验，通过PRIMEPOWER分区为它的PRIMEQUEST Itanium服务器提供了独特的性能。

分区粒度

正如前面所描述的，PRIMEQUEST系统是由模块化构件构成的。每个系统板最多可包含4个处理器芯片、256GB内存，通过系统互连crossbar方式连接到I/O。通过给crossbar交换机下发指令，把系统板集合与其他的系统板隔离开来，并给这些系统板分配I/O单元访问权限，就可以在PRIMEQUEST中创建独立的分区。每个独立的分区都运行一个不同的操作系统实例，都与其他分区中的硬件或软件故障之间实施电隔离。

图 2: PRIMEQUEST灵活的分区



重新配置系统板分区并不是PRIMEQUEST系统所独有的。其他高端服务器平台也可能具有类似的性能。但是，富士通的PRIMEQUEST服务器的实现方式是独立于系统板分配I/O单元。其他设计有时候包括系统板或套装系统板与I/O之间的I/O连接，这些连接在某种程度上限制了指派不同的系统与I/O板组合的灵活性。PRIMEQUEST灵活的I/O设计允许将I/O单元与任何系统板关联起来（通过crossbar交换机），允许通过适当地组合处理能力和I/O连通性来规定分区。

图2说明了将一个具有八个系统板的PRIMEQUEST服务器划分成四个分区。每个分区都运行着它自身的操作系统（Linux或Windows），其中包括不同的软件版本或不同的软件栈。图2还说明了可以灵活地分配I/O单元：一个分区具有两个系统板和三个I/O单元，而另一个分区具有四个系统板和两个I/O单元。

图3. 利用XPAR功能分区

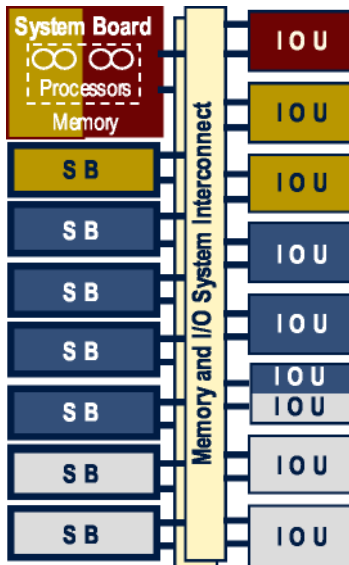


图3阐明了PRIMEQUEST XPAR提供的扩展分区。XPAR功能允许将一个系统板分成2个分区，每个分区都具有2个双核处理器芯片或者总共具有4个内核。类似地，I/O也可以分割，分配给不同的分区。利用XPAR，图3顶部的分区由具有一个I/O单元的半个系统板（2个芯片，4个内核）组成。下一个分区使用一个半系统板（总共具有6个芯片和12个内核）和2个I/O单元。其它两个分区共享1个I/O单元。XPAR提供的较小的粒度呈现了某种程度的分区灵活性，这是其它基于IPF的服务器所不具备的。

非镜像运转和多种镜像运转都可以使用分区。但出于工程方面的考虑，混合多种模式存在一些限制。如果系统处于非镜像模式，那么所有的分区就必须都是非镜像的。如果系统被配置为镜像的，那么每个分区就必须以一种镜像模式运行：标准镜像(Standard Mirror)或扩展镜像(Extended Mirror)。由于采用标准镜像不存在性能或成本损失，所以假定客户运行系统时会选择启用镜像模式。需要扩展镜像的额外可用性的应用软件可以将分区设置为以那种模式运行，而其他分区仍然以标准镜像模式运行。

并发维护性能

硬件分区上的强制电隔离确保一个分区的故障不会影响其他分区。如果一个处理器或内存模块发生故障并且处于脱机状态，剩下的处理器和内存仍然能继续运转。但为了更换出现故障的芯片，需要取下整个系统板。类似地，如果PCI板卡出现故障，也能进行更换，但如果故障出现在I/O单元电路中，就必须取走I/O单元进行维修。

关于发生故障时可以更换组件而不会影响整个系统的其他部分的具体细节，请参看集中讨论可用性的一份姐妹篇白皮书。一般而言，如果一个未被镜像的组件出现故障，那么使用该组件的所有分区就都会停止。故障可以在带电状态下热插拔，以便未受影响的分区能够继续运转。如果出现故障的组件通过一个冗余组件进行了镜像，那么受影响的分区就能继续运转，可以推迟维修。可以在方便的时候停止受影响的分区，取出热插拔组件进行维修。表4（下面）总结了一些可以同时维修的主要组件。更多细节，请参考姐妹篇可用性(Availability)白皮书或者PRIMEQUEST文档。

内存	在可选系统镜像中，内存为冗余配置，不会中断分区。在标准系统镜像中，分区会失效，但可以重启，从而忽略发生故障的内存DIMM。修复需要系统板的热插拔。
系统板	更换系统板时，使用该系统板的所有分区必须停止运行。支持热插拔，使其他分区可以继续运行。
地址 Crossbar	标准系统镜像的地址Crossbar为冗余配置，如果一个组件失效，会自动实施隔离。隔离状态下，发生故障的组件可以热插拔。需要系统重启以重新激活修复的组件。
数据Crossbar	可选系统镜像的数据Crossbar为冗余配置，如果一个组件失效，会自动实施隔离。隔离状态下，发生故障的组件可以热插拔。需要系统重启以重新激活修复的组件。
I/O组件	如果发生故障，分区停止运行。分区停止时，发生故障的I/O组件可以热插拔。其他的I/O组件可继续运行。
系统管理板	可冗余配置。如果活动的系统管理板出现故障，系统自动切换到备份板上。支持热插拔。
KVM交换机	非冗余配置，但不会引起系统崩溃。支持热插拔。
G比特以太网集成式交换机	出现故障会影响连接的LAN的服务。如果分区配置为冗余的GbE连接，分区可以继续运行，改变LAN连接。
风扇盘	如果为冗余配置，风扇盘失效时，系统仍可以继续运行。
电源	支持N+1电源。一个电源失效时，系统仍可以继续运行。支持热插拔。

表4： PRIMEQUEST并发维修概要

IDEAS概要

富士通的PRIMEQUEST关键任务IPF服务器的设计目标在于，满足客户对于采用业界标准软硬件的企业级服务器的需求。由于最多可扩展至64个Intel Itanium 2处理器和2TB内存，所以它能满足客户对于高性能、共享内存的多处理器的需求。当然，并非所有的关键任务应用都需要一个完整的大型SMP。因此，PRIMEQUEST服务器提供了灵活的基于硬件的分区特性，可以创建多个隔离的分区。

富士通的设计在所有基于Itanium的SMP中独树一帜，原因在于极其关注硬件的高可用性。除了广泛使用ECC和其他传统的可用性技术之外，独特的镜像crossbar和镜像内存向最苛刻的客户保证，PRIMEQUEST平台能用于满足关键任务计算。客户可以基于每个分区选择不同的镜像级别，以满足他们的工作负载需求。

从硬件角度来看，富士通开发的Itanium服务器的确令人印象深刻。Ideas International (IDEAS) 认为，客户会认识到富士通致力于设计出适合于关键任务计算的机器，会乐意将PRIMEQUEST服务器看作是昂贵的专有系统的一个竞争者。富士通合作伙伴的即时支持，尤其是PRIMEQUEST独有的关键任务特征的支持，将在确保这种新型计算平台的市场接受程度方面发挥重要作用。

尾注

本白皮书撰写期间，根据业界标准基准度量的性能评测结果还未出现。最新的关于采用各类处理器的服务器性能比较排行榜，请参看：<http://www.ideasinternational.com/benchmark/bench.html>

Single-4bit block Error Correction – Double-4bit block Error Detection

参阅Ideas International于2006年7月发布的姐妹篇白皮书“富士通新的高端双核Itanium PRIMEQUEST 服务器提供最大的高可用性”。

Americas

Ideas International, Inc.
800 Westchester Avenue
Suite S620

Rye Brook, NY 10573-1330
USATel + 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

