



PRIMEQUEST 概览

2006年7月

为Fujitsu编制

TABLE OF CONTENTS

摘要	1
引言	2
业界标准的关键任务高端服务器 拥有巨大的客户需求	2
PRIMEQUEST家族	3
PRIMEQUEST 500系列	3
结构概览	3
关键任务设计的领先性	4
镜像总线、存储器	4
操作系统支持	5
继续开发	5
高可用性功能	6
热插拔	6
高可用性集群	7
系统管理	7
系统管理软件	7
系统管理硬件	8
系统管理接口	8
IDEAS概要	9
尾注	10

TO DIG DEEPER

本白皮书由Ideas国际 (IDEAS)编制,对富士通的PRIMEQUEST服务器进行了概述。四本姐妹篇白皮书在本概览基础上进一步详细介绍了以下方面:系统架构、高可用性、操作系统、以及系统管理。

摘要

客户正在不断寻求“开放的”硬件和软件以最小化对私有方案的依赖性。2002年,富士通制定了开发关键任务平台的策略,该平台基于开放标准构建,创建企业级服务器,利用业界广泛接受的标准Intel处理器及运行在该处理器上的Linux和Windows操作环境。2005年,富士通发布了PRIMEQUEST 400系列的第一款关键任务IA服务器。随着双核Intel Itanium2处理器的上市,富士通更新了基于Itanium的企业服务器产品线。双核处理器加上富士通独家提供的新特性,使PRIMEQUEST 500系列具有更强的可扩展性。

富士通服务器与其他基于Itanium处理器族(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UNIX PRIMEPOWER的设计经验,创建出基于Itanium的服务器系列,适用于关键任务负载的运行。PRIMEQUEST系统是基于工业标准的服务器,其独有特点是可提供关键组件镜像的多种方案,在发生硬件故障时能够保证连续运行。系统具有冗余的地址和数据交叉开关,因此可传送双重信号。内存也可冗余配置,在不同模块对数据作镜像。

PRIMEQUEST也提供灵活的分区功能,最多可配置8个隔离分区,每个分区运行自有的操作系统实例和软件栈。富士通实现的独特之处在于其灵活的I/O功能可将系统板和I/O单元独立分配给分区,从而根据最适合的处理能力和I/O连接性组合为应用集进行分区定义。

总之,富士通利用其设计企业级服务器的丰富经验,开发出PRIMEQUEST服务器。即使是对以前富士通高端系统不熟悉的用户,也会认可PRIMEQUEST的独特功能使其成为一个极具吸引力的平台,是适用于关键任务应用的业界标准服务器。本文正文部分简要介绍PRIMEQUEST服务器设计的关键架构特点,并论述最近发布的采用了双核Intel Itanium2处理器的产品(以前的产品代号为“Montecito”)。相关白皮书则提供了附加的细节,诸如系统架构、操作系统支持、高可用性功能及系统管理功能。

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、或授予第三方分许可。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。本白皮书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。

引言

业界标准的关键任务高端服务器拥有巨大的客户需求

50多年以来，用户依靠大型富士通计算机来解决重要的业务问题。随着产业的发展，富士通创造出领先的产品，满足用户独特的需求。凭借大型机和超级计算机的设计经验，富士通为寻求Solaris操作环境企业级性能和可靠性的用户开发出引人注目的PRIMEPOWER家族UNIX服务器。这些基于SPARC64的PRIMEPOWER服务器仍然是富士通产品线的一个重要部分。

现在，很多客户对Linux操作环境和采用IPF芯片的服务器很感兴趣。传统UNIX操作系统，如AIX、HP-UX或Solaris十分成熟且功能齐全。但是每个系统都由其支持厂商开发，因此它们都是私有的。此外，并不是所有ISV（独立软件供应商）应用都在所有私有UNIX系统上可用。历史上每个私有UNIX系统都进入了不同的工业领域。这就是说，在某个领域可能普遍使用一种UNIX，而在其他领域却完全没有这种UNIX的用武之地。ISV一般集中资源支持满足最大需求的环境。因此专用ISV应用通常并没有移植到所有私有UNIX变体中，或者在每种UNIX上支持的程度不同。

许多客户也关心受限于固定的供应商或缺少板卡的应用可用性。他们重视以Linux为代表的开放、协作的环境。政府机构尤其欢迎Linux，希望可以保证软件投资的长期性，无需担心供应商产品的稳定性及其扩展操作系统投资的能力。Linux代码开放，经过充分测试，这给客户带来安全感，认为Linux是安全稳定的环境。此外，在各个领域部署的Linux越来越多，ISV也把Linux作为主要软件平台，支持各种应用。当然，Linux的低成本也十分具有吸引力。然而，仅仅低成本是不足以满足关键任务方案的。Linux不仅成本低，而且还可以提供不受供应商限制的稳定、灵活的操作环境。

企业客户对Microsoft Windows也越来越感兴趣。Windows虽然不像Linux那样在开放的合作环境中开发，但Windows的广泛使用能保证缺陷和问题迅速定位并修改。此外，大量的用户群也吸引很多ISV支持各种应用。

偏好Linux和Windows的客户同样都希望服务器采用业界标准的处理器。处理器设计和芯片制造需要大量投资才能保持竞争力。客户担心会受限于少数系统供应商的私有处理器。由于这些供应商往往专注于其自身的强势领域，因此他们的平台只在相对较少的领域适用，这就降低了ISV支持的可用性。另一方面业界标准的处理器，从其定义就可看出，能支持各种工业领域的产品。Intel的高端IPF产品线，尤其是Itanium2实现，可提供企业计算所需的高性能处理器。

客户欢迎Linux的开放性、Windows的广泛性以及选择业界标准处理器构成的各种平台的机会。然而正是这种选择给供应商带来了挑战——在遵守业界标准时如何对系统做出区分？富士通根据其关键任务服务器长期的设计经验做到这一点。本白皮书，以及本系列的其他白皮书，介绍了PRIMEQUEST产品区别于其他支持Linux和Windows的平台的特点。

PRIMEQUEST家族

PRIMEQUEST 500系列

2004年发布的第一个 PRIMEQUEST 产品采用单核 Intel Itanium2 处理器，可容纳多达9MB的片内高速缓存。400系列产品有支持8、16或32个Itanium2 处理器芯片的不同配置。随着双核 Itanium2 处理器芯片的发布，富士通推出了PRIMEQUEST 500系列。这些新产品也有支持8、16或32个Itanium2 处理器芯片的不同配置，但得益于双核处理器，同以前的型号相比，现在容纳了两倍的内核。

正如在系统架构白皮书中所论述的，双核Intel Itanium2 处理器芯片也包含更大的片内高速缓存和支持芯片组的更快总线。加上富士通设计了更快的交叉互连，采用双核 Intel Itanium2 处理器的500系列同类似的单核400系列产品相比，可提供两倍的吞吐量。

Intel正在开发更强大的双核处理器，预计2007年下半年推出（虽然具体细节尚不清楚）。为了满足客户对投资保护的要求，富士通的PRIMEQUEST系统精心设计，能适应未来的多核 Itanium2 处理器。

结构概览

如同其他可扩展的SMP设计一样，PRIMEQUEST系统也通过模块化的建构模块制造。系统板包括处理器和内存。PRIMEQUEST系统板有四个处理器芯片和32个内存DIMM插槽。I/O单元提供PCI插槽和千兆以太网以及SCSI连接。交叉开关提供处理器/内存板卡和I/O单元的系统互连。富士通设计的系统管理板卡监控故障状态，配置分区中的系统。

仅从基本描述来看，PRIMEQUEST结构与其他模块化可扩展SMP系统的高级描述十分相似。但有一些重要的不同之处。比如，系统互连（交叉开关）是整体性能的关键决定因素。富士通利用其超级计算的开发经验设计交叉开关，以较低时延提供更高带宽。尤其是PRIMEQUEST 580可提供137 GB/秒的汇聚峰值带宽。这样，即使采用相同的处理器芯片，PRIMEQUEST也能提供比其他实现更好的系统性能。

富士通系统互连不仅能提供高性能，也可以配置成冗余互连，每个互连传送相同数据的双倍拷贝。一组硬件发生故障时，PRIMEQUEST系统可不中断业务继续运行。系统互连的冗余镜像是确保PRIMEQUEST在关键任务环境中工作的唯一镜像选择。后续章节中有关于富士通独有镜像功能的更多讨论。

另外一点就是系统板和I/O单元可独立地分配给不同分区。其他设计也许能把I/O功能和相应的系统板联系在一起，但需要把系统板和I/O始终配置在一起。而富士通的方案则不需要预先设计哪个I/O单元与特定的系统板联系，这样各个分区就可灵活地包含不同数量的I/O单元。PRIMEQUEST分区特点将在本文的后面部分进一步进行讨论。

关键任务设计的领先性

镜像总线、存储器

尽管富士通已经在高度可靠的Intel处理器上采用了传统的错误校验和纠错(ECC)技术，但有的关键任务客户环境还需要更高的可用性。针对这种情况，富士通提供了一系列其他高端服务器无法匹敌的镜像选择。通过采用富士通大型机技术，系统镜像操作可复制(硬件冗余)分区资源(交叉开关、存储器)并在锁定步骤下运行。因此，这些组件的任何故障均不会导致分区故障。

特别是，PRIMEQUEST提供对交叉开关和内存处理器进行镜像的各种方案，确保故障发生时操作继续运行。PRIMEQUEST提供两种镜像模式——标准模式和扩展模式，此外还可以运行在非镜像配置选项下。

镜像配置时，资源成对配置，并行地执行相同功能。一个资源发生故障时，其他资源仍可提供有效操作，系统可不中断运行。标准镜像功能只对全局地址交叉开关做镜像，作为系统板和I/O单元之间的地址互连。扩展镜像把四个全局数据交叉开关安排为2个重复的组件并配置内存，因此在两个分隔的内存模块中所有数据都有重复位置。

PRIMEQUEST服务器的默认镜像是标准镜像。在这种模式下地址交叉开关分成两部分，每一部分都有相同数据的拷贝。由于地址交叉开关上的负载不大，因此将其分成两个镜像拷贝不会给整体系统性能带来显著影响。

扩展镜像模式则更进一步，复制全局数据交叉开关以及镜像内存。对数据交叉开关做镜像后，系统板接口电路也在镜像配置中运行，以此保护接口及交叉开关本身的系统故障。扩展镜像也复制存储在主存中的数据。存储器由ECC代码保护，不仅可纠正单比特错误，而且可重构缺陷内存芯片的所有四个比特。然而如果更

多错误发生，则存储器在该位置的内容无法恢复。内存镜像在两个独立的内存芯片位置上复制数据。可能会使内存崩溃的多比特故障基本不可能同时在镜像位置发生。此外，接口电路的镜像也避免了电路故障带来的系统故障。因此，镜像内存可提供关键任务应用所需稳定的数据完整性。当然，镜像内存确实会把整体内存容量减半。PRIMEQUEST服务器可配置扩展内存，因此总体内存容量的减少不是问题。不过把数据存储两个内存中确实会使内存成本加倍。因此，扩展镜像会带来经济上双倍的成本。至于性能影响，富士通的数据交叉开关镜像设计确实会减小可用交叉开关带宽，但性能影响不会超过5%。

操作系统支持

2006年9月推出的一般可用性上，PRIMEQUEST 系列500服务器支持：

- » Red Hat企业级Linux 第四版(RHEL4)。稍后会支持更新4和5 以及RHEL5（随Red Hat的产品发布时间而定）。
- » SUSE Linux企业级服务器第9版（SLES9），以及SLES10（只待它一发布，这取决于Novell）。
- » Windows Server 2003，面向Itanium并整合了SP2的企业版（多达8个处理器芯片/16个内核以及最大256GB内存），将于2007年初发布（取决于Microsoft）。
- » Windows Server 2003，面向Itanium并整合了SP2的数据处理中心版（多达32个处理器芯片/64个内核以及最大2TB内存），将于2007年初发布（取决于Microsoft）。

继续开发

在PRIMEQUEST服务器中Windows和Linux均可使用，且无需更改。然而富士通正与Microsoft和开源社区紧密合作，为Windows和Linux添加新功能，使其更适合于PRIMEQUEST服务器以及其他计算机系统。

比如在Windows方面，富士通正与Microsoft合作，PRIMEQUEST服务器上使用的Vista中将集成动态重构功能。

在Linux方面，富士通正和开源开发实验室（OSDL）、Red Hat和Novell SUSE合作，在其各自的Linux发布及普遍的Linux社区中添加功能。这些功能将解决dump能力、高可用性集群、可扩展性、RAS（可靠性、可用性和适用性）、资源管理、动态重构以及可视化问题。现在这些领域的一些功能已经在RHEL4和SLES9（Linux内核2.6）中可用。其他将在RHEL5和SLES10中可用。

有关操作系统支持的更多细节（现有的或计划中的）可在操作系统白皮书中找到。

高可用性功能

先进的Itanium处理器设计和PRIMEQUEST系统功能（如可镜像的总线与存储器），为高可用性的关键任务计算提供了坚实的基础。以下所突出的其它功能（在高可用性白皮书中有完整描述）进一步增强了PRIMEQUEST的高可用性特点。

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、键盘/视频/鼠标（KVM）单元，以及其他系统组件。

关于哪些可以热插拔的详细分类取决于特定的PRIMEQUEST型号以及PRIMEQUEST客户所选取的冗余选项。许多主要组件不能热插拔，除非把它们从运行的分区中卸载下来。这样的组件包括系统和I/O单元，及其所包含的组件（处理器、内存等等）。不在运行分区中的系统和I/O单元无需动态重构就能替换。替换完成后运行的分区不受影响。单独的系统板处理器芯片和DIMMS只有在把系统板从PRIMEQUEST底板上拆下后才能替换。

在Linux和Windows支持动态重构功能后，硬件热插拔功能就更容易实现。比如使用动态重构后，系统和I/O单元可以热交换，某些交叉开关（如运行分区中的数据交叉开关）可以先从服务中断开，进行修复，然后使其重新工作，并不会带来应用服务的中断。然而动态重构取决于冗余路径（如I/O）、分区中的系统板编号等等。高可用性白皮书（参考较早的）简要介绍了一些热交换细节。更多的具体信息可参考富士通的PRIMEQUEST文档。

高可用性集群

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。

另外，单一PRIMEQUEST机架可配置成内部集群结构（如工作分区和备份分区）。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，为重启应用服务提供故障切换机制。而镜像意识不到应用服务的中断，无法发现应用故障并且没有故障切换机制。

富士通 PRIMECLUSTER 产品为 PRIMEQUEST 系统提供必要的集群软件。PRIMECLUSTER 是工作在 Solaris 上经过长期检验的成熟产品，也可在 Linux 上工作。

请注意当 PRIMEQUEST 分区运行在任何一种镜像模式下时也可配置集群。可进行基于光纤通道的共享磁盘配置和多种其他的集群 I/O 配置。

系统管理

系统管理软件

富士通为客户管理 PRIMEQUEST 单机服务器或集群服务器提供多种选择。大多数 PRIMEQUEST 客户将更愿意把 PRIMEQUEST 系统管理板当作最实用、最强大和最具有成本效益的系统管理技术。这种板能提供一个 PRIMEQUEST 服务器安装所需的管理功能，可望成为最通用的管理技术。系统管理板已成功应用于较早的 PRIMEQUEST 400 和 PRIMEQUEST 500 系列的第一个产品上。

一般来说，对于安装了 PRIMERGY 和/或 PRIMEPOWER 的客户，富士通也提供 Server View 和/或服务器管理器系统管理软件。这些软件可同时管理多个富士通系统，并且无需基于相同的操作系统。ISV 提供的企业系统管理 (ESM) 软件可作为替代选择。多数情况下运行在 PRIMEQUEST 系统中的代理软件能够向 ESM 控制台提供所需管理信息。

日本客户常选择富士通的 Systemwalker 作为 ESM 软件。它提供分区、机架和电源管理，以及 PRIMEQUEST 服务器的报警信息。Systemwalker 也使用 SNMP 协议来接收其 PRIMEQUEST 服务器信息。机架信息从 Systemwalker 连接的 PRIMEQUEST 系统管理板中得到。

系统管理白皮书探讨了这些其它的选择，并提供有关 PRIMEQUEST 系统管理板功能的更多细节。

系统管理硬件

系统管理板是 PRIMEQUEST 关键任务能力的一个主要因素，为单个 PRIMEQUEST 服务器提供管理信息。也提供 API，从而实现富士通服务器管理器或其他第三方企业管理系统对 PRIMEQUEST 系统的管理。

系统管理板控制台（通常是一台PC）提供基于浏览器的安全连接，通过PRIMEQUEST管理LAN，连接PRIMEQUEST系统管理板。这种系统管理选项提供机架的硬件信息、分区配置信息、操作系统、维护和资源管理信息及控制。

系统管理板也监视电源控制、电压电平和组件温度。此外，系统管理板也控制在单个PRIMEQUEST分区中使用标准镜像还是扩展镜像。最后，系统管理板还用来初始化和关闭系统。

所有PRIMEQUEST产品都提供系统管理板。对于540和580（以及440和480）这几款产品，可配备第二个系统管理板以提供冗余能力。

系统管理接口

系统管理板的功能是为各种PRIMEQUEST系统组件提供统一管理，包括硬件、软件、分区和机架。系统管理板通过独特结构实现该功能。

比如，某些UNIX服务器通过LAN连接到外部管理服务器，与此不同的是，PRIMEQUEST服务器本身具有内建的PRIMEQUEST系统管理板。这种结构无需外部服务器、外部LAN连接和web服务器软件。需要的一切只是一个浏览器，这就极大地简化了管理负担，增强了安全性。

为了完成所有功能，系统管理板采用微处理器服务、相关电路、以及富士通具有高可用性独特设计的ASIC。同时也有管理LAN hub和专用LAN hub。这些hub完成系统管理板和PRIMEQUEST组件、PRIMEQUEST分区和外部世界的通讯。

系统管理板及其到PRIMEQUEST服务器和外部世界的连接高度可用。第一级高可用性由两个按主-从模式配置的系统管理板决定（420或520型不提供冗余配置）。如果一个发生故障，则另外一个接管任务执行。此外，富士通时刻关注高可用性，因此系统管理板和逻辑系统板/分区也通过冗余的100MB/秒的以太网LAN连接。

如果系统管理板有安全缺口，则扩展管理功能的高可用性对客户就毫无价值可言。为了使这种情况发生的可能性降至最小，系统管理板和系统管理控制台之间的信息交换由具有128位加密编码的SSL（安全套接层）通信保护。

为了进一步增强安全性，系统管理板使用IP过滤来限制访问。在四个用户帐户级别对访问加以控制——管理员、操作人员、用户和服务人员。此外，系统管理板保存的日志可显示哪些人员在何时处于何种目的访问板卡。

IDEAS 概要

建立在富士通长期大型机、超级计算机和UNIX处理器设计经验之上的PRIMEQUEST服务器，采用了业界标准的软硬件，满足客户对企业级服务器和关键任务服务器的需求。采用业界标准的硬件（Intel IPF）和操作系统（Windows和Linux），PRIMEQUEST可直接在世界市场上与来自HP、SGI、IBM、NEC和Unisys的机器竞争。美国市场的主要目标是HP、IBM和Unisys。目标应用包括OLTP、决策支持/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。PRIMEQUEST服务器将在所有富士通服务器产品线中出现，补充UNIX (Solaris)的PRIMEPOWER系统，以及PRIMERGY服务器（运行Linux和Windows的业界标准的Intel Xeon处理器）。

由于富士通专注于硬件的高度可用性，因此PRIMEQUEST设计不同于其他基于IPF的系统。除了大量使用ECC和其他传统可用性技术外，PRIMEQUEST系统独特的镜像互连交叉开关和镜像存储器使得要求最高的客户也能确信该服务器可用于关键任务的商业计算。

从硬件角度来看，富士通确实创建了一款不平常的基于IPF的服务器。客户会注意到富士通如何专心设计适合关键任务计算的机器。富士通合作伙伴的及时支持，尤其是对PRIMEQUEST独特的关键任务特性的支持，将在确保该计算平台的市场接受度上起到关键作用。目前在PRIMEQUEST服务器上可使用Windows和Linux发布，而无需作任何更改。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系统。

Ideas国际（IDEAS）确信客户将乐意把PRIMEQUEST系统作为关键任务计算的选择。可能只有很少应用要求镜像功能的全部集合（包括双重存储器）。然而对于这类环境来说，PRIMEQUEST平台是替代昂贵的私有系统（如大型机）的唯一业界标准的选择。其他客户可按每个分区基础选择不同的镜像级别以满足工作要求。对于这些客户，PRIMEQUEST平台使用业界标准达到的高性能与其高可用性具有同样的重要价值。

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

**尾注**

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 “富士通基于Itanium 2的新型高端PRIMEQUEST服务器为Linux和Windows应用提供关键业务托管”， Ideas国际,2006年7月。

参看相关白皮书 “富士通基于Intel Itanium 2的新型高端双核PRIMEQUEST 服务器继续提供业界领先的系统管理”， Ideas国际，2006年7月。