

# 安腾技术文集(修订版)



# HP 基于 IPF 的 Integrity 系列 开创了工业标准 64 位应用的新时代

## 一、HP 基于 IPF 的 Integrity 服务器系列概述

### 1.1 64 位服务器技术新发展

2003 年中 HP 宣布推出基于 Itanium2 的 Integrity 系列，成为业界第一个全面覆盖各个档次应用的服务器系列，把服务器技术推向新的高度，在 IT 技术应用史上揭开了基于工业标准的 64 位计算的新篇章。

#### 1.1.1 HP 推出基于 IPF 的 Integrity 服务器系列

从 1992 年基于 Alpha 的服务器问世开始，以当时的 DEC 公司为代表 64 位 RISC 服务器厂商逐步完成了从 32 位过渡到 64 位技术各项任务，充分展示了 64 位技术的优越性，计算机技术高端应用经历了长达 10 年的 64 位 RISC 时代。从 2001 年 HP 等厂商推出基于工业标准的 IPF 处理器系列第一代 Itanium 的服务器和工作站产品开始、到 2003 年 HP 宣布推出基于 IPF 处理器第二代 Itanium2 的 Integrity 系列，HP 以短短的 2 年时间完成了从 64 位 RISC 处理器发展到 64 位 IPF 的各项任务，包括提供从入口级到企业级完整的服务器系列(以及工作站产品)、新系统平台上的 64 位操作系统、系统基础件和业务基础件平台、丰富的解决方案和应用软件、服务和支持，建立了完整的生态系统以及基于它的一整套系统开发和应用技术，创造了大量成功的应用实例、应用经验和应用成果，使人们看到了工业标准 64 位应用新时代的曙光。从此，服务器高端市场的竞争形成了 64 位 RISC 和 64 位工业标准 IPF 两军对持的新态势。虽然基于 IPF 服务器产品要夺取 64 位 RISC 主要市场份额还需要假以时日，但是以 HP Integrity 系列问世为标志，人们越来越清楚地认识到工业标准的 IPF 将代表未来的发展方向、具有不可限量的发展前景。

#### 1.1.2 服务器产品的分类

当前服务器市场主要是基于 Intel Itanium 处理器系列和 64 位 RISC 产品的竞争。此外，不少厂商还提供基于工业标准 IA-32 的 Xeon 或 Xeon MP 32 位服务器，也在中低端占有相当大的市场份额。所有服务器产品根据其装备的 CPU 个数、价位和应用可以分为入口级、中档和企业级等三个档次。

#### 按照服务器装备的 CPU 个数分类

考虑到当前各厂商用于装备服务器的 CPU 主要可分为三类：基于工业标准的 Intel Itanium 体系结构处理器、64 位 RISC 处理器和工业标准的 IA-32 处理器。32 位处理器性能显然远低于同样数量的 64 位处理器，64 位 RISC 处理器也大致上可以分为快慢两类：美国著名的咨询公司 Illuminate 在对所有 RISC 微处理器以及工业标准的 Intel IA-32 微处理器性能指标发展和变化进行了长达 4 年的跟踪和比较后，在 1999 年发表调研报告指出：“在今天所有 RISC 体系结构处理器中，只有康柏的 Alpha 和 HP 的 PA-RISC 在通用计算中具有超级的性能。相

反, IBM 的 Power、SGI 的 MIPS 和 Sun 的 SPARC 性能都比较落后, 甚至低于业界公认的基线 Intel IA-32。”(注: 目前 IBM 基于多核技术的 Power4 处理器也能够提供与 Alpha 相竞争的性能, 我们也将它归入高性能 RISC 处理器一类。)大量基准测试的结果表明 Itanium2 是当前速度最快的 64 位处理器, 装备 Itanium2 的服务器的性能指标相当于或者高于装备 2 倍数量 SGI /Sun 的 R12000 /UltraSPARC III 的服务器的性能。因此, 我们将按照系统装备高性能的 Itanium2 或 Alpha/PA-RISC/Power4 处理器的个数来分类服务器产品, 而把装备 2 倍数量其他 64 位 RISC 或 IA-32 处理器的服务器产品归入同一类型:

- 入口级服务器: 装备 1-4 个高性能 CPU 的服务器产品;
- 中档服务器: 装备 8-16 个高性能 CPU 的服务器产品;
- 企业级服务器: 装备 16 个以上高性能 CPU 的服务器产品;

## 按照服务器价位分类

当前, 由于存储设备和网络设备作用越来越重要, 存储系统和网络通信设备规模和水平在企业 IT 项目投资中所占比例越来越大, 服务器总体配置规模的往往成为投资的决定性因素(如磁盘容量、是否采用网络存储系统、是否装备高速打印机或高分辨率图象/图形显示和硬复制设备等)以及内存容量等, 而不是 CPU 数量。但是, 服务器系统的配置是与装备的 CPU 个数紧密相关的, 即 CPU 数少的服务器一般不会(甚至也不允许)装备很大的内存和磁盘容量或其他高档外设。因此, 服务器产品也可以按照价位来分类(这样的分类一般不会与按照 CPU 分类产生很大的矛盾)。下面我们根据 IDC 2002 年发表的报告中使用的分类原则来划分市场上服务器产品的类型:

- **入口级服务器:** 价格低于 10 万美元 (装备 1-4 个 CPU 的服务器价格不会高于 10 万美元);
- **中档服务器:** 价格低于 100 万美元(事实上, 有些装备 8 个 CPU 的服务器系统价格甚至低于 10 万美元, 这里给出的是价格上限);
- **企业级服务器:** 价格低于 300 万美元(价格超过 300 万美元的计算机系统一般将被列入超级计算机系统范畴);

## 按照应用分类

从应用角度虽然很难将服务器产品严格分类, 但是各个档次的服务器产品也都有其主要的应用领域, 由此也决定其系统资源配置、RAS 和可管理特性和价格, 即服务器的基本特征:

- **入口级服务器:** 在当前流行的 3 层结构企业应用在入口级服务器主要应用于企业网络边缘或工作组服务器, 也可以作为用户的桌面系统; 在企业网络边缘, 入口级服务器主要作为网络安全、防火墙或者通信服务器; 在桌面上, 入口级服务器可以作为更高档次服务器的应用软件开发机使用; 为了支持工作组应用, 入口级服务器可以提供文件存储浮点计算和打印等方面服务; 在通信、电信和网络服务供应商(ISP)等行业, 入口级服务器被大量地装在机架上, 基本上运行固定的软件, 以其高于 IA-32 服务器的寻址空间、性能和可用性, 支持用户单位的主要业务; 入口级服务器可以组成 UNIX 或 Linux 集群支持

高端应用等等。入口级服务器一般配置较少的外围设备，有时甚至象一台普通的仪器，放置在工作地点、默默地发挥作用。当前，入口级 RISC 服务器在性能和价格/性能方面受到性能越来越高、价格/性能更好的 IA-32 服务器的挤压，但由于它们能够运行与中、高端 RISC 服务器相同的 UNIX 操作系统和软件兼容以及 64 位的特性、价格也越来越便宜，入口级 RISC 服务器仍然有其生存的空间。基于 IPF 的入口级服务器以其高性能和工业标准的批量优势与开放源代码 Linux 结合，成为这一领域中一匹强有力的黑马。今后各厂商基于上述三大类处理器的入口级服务器产品将继续展开剧烈的竞争；

- **中档服务器：** 在当前流行的 3 层结构企业应用中，中档服务器主要用作第二层的应用服务器或者后端的数据库服务器，也可以作为用户的桌面系统使用。中档服务器广泛应用于企业和高性能技术领域，作为大容量数据库服务器、高性能计算应用服务器、网络文件系统服务器或 Internet 服务器。中档服务器一般有两大类应用形态：一种是前台应用，用户直接在其上运行大型应用软件，用户可以把它作为自己工作环境中的一台高性能的计算机运行科学计算、数据库管理、企业资源规划、客户关系管理、供应链管理等方面大型软件；另一种是作为中、小企业和大型企业部门的服务器，在一定范围内提供后台支持、发挥网上应用服务器的作用。中档 UNIX 服务器是应用最广泛最广的 RISC 服务器，它的应用范围几乎可以覆盖计算机的整个应用领域。无论哪种应用形态，中档服务器一般需要配置一定规模的系统资源如较大的内存和磁盘容量，提供较高的 RAS 和管理特性，方能完成自身所承担的任务。近年来，基于工业标准处理器的服务器在中档应用和市场发挥越来越大的作用：基于 IPF 的中档服务器以更高的性能和性价比成为这一领域有力的竞争者；基于 IA-32 的中档服务器也可以与 64 位系统相比拟的性能和更高的性价比，成为许多企业特别是中小企业的选择对象。三者今后将继续剧烈竞争；
- **企业级服务器：** 企业级服务器拥有极其强大的信息处理能力和丰富的系统资源，一般都放置在各种企业、研究院所、大学、政府和军事部门的信息中心、数据中心、客户服务中心、交易中心、管理中心等等，通过网络对遍布在很大范围内的用户提供服务。随着 Internet 和电子商务的爆炸性发展，许多大型的电子商务企业和 Internet 门户网站提供全球服务的功能。高档 RISC 服务器一般有两大类应用要求：一类是能够在网（特别是 Internet 网）上支持大量同时用户，具有此类需求的应用非常多。例如，股市交易中心、电子商务中心、各种大型服务中心、呼叫中心等等；另一类是能够快速解决超大规模的计算问题如全球气象预报、地震预报、大规模的模拟和仿真应用等，或者能够管理超大规模的数据库或数据仓库等，提供很强的商务智能。无论哪类应用，都要求企业级服务器能够扩展到提供很大的规模的系统资源，具有支持连续运行的高可用性。企业级服务器台数虽然比中、低档服务器少，但是重要性却大得多。考虑到各种连带的业务如用户往往倾向于购买与企业级服务器相同类型的中档、入口级服务器、工作站以及技术服务，企业级服务器已经成为服务器市场竞争的焦点。从某种意义上可以说：“得企业级服务器者得天下”，也就是说哪个厂商能够占领企业级服务器的制高点，哪个厂商就将在竞争中获胜。当前，IA-32 服务器还比较难进入企业级市场，HP Integrity 系列 Superdome 服务器能够扩展到支持 128 个高速的 Itanium2 处理器，提供超过 64 位 RISC 服务器的性能和性价比，推动基于工业标准的 IA-64 处理器的服务器产品进入企业应用的最高端，打破了 64 位的 RISC 的一统天下，使今后高端应用将成为 IPF 与 RISC 的战场；

必须指出的是服务器产品的分类并不是严格的，不少产品既可归入入口级也可归入中档或者既可归入中档也可归入企业级。当然，价格贵、配置小、应用层次低的产品在市场上肯定是站不住脚、肯定将遭到淘汰。

### 1.1.3 HP Integrity 系列概貌

HP 提供业界最广泛的服务器产品阵容，全面满足用户的需求。其中，HP Integrity 系列是当前唯一的覆盖各个档次应用的基于工业标准 IPF 处理器服务器系列，提供领先的入口级、中档和企业级服务器，奠定了服务器应用进入 64 位工业标准新时代的基础。

其中，基于 Itanium2 的 Integrity 系列的概貌如表 1-1 所示。

	ProLiant	Integrity	HP9000	AlphaServer	NonStop
数据库					
应用	 BL p-Class	 rx7620	 rp8400	 rp7410	 ES series
访问	 BL e-Class	 rx4640  rx2600	 rp5430/70  rp2430/70		 S76 series
HP 提供业界最广泛的基于 Itanium 服务器产品系列!					

图 1-1 HP 提供世界上最宏伟、最牢固的服务器产品阵容

### 1.1.4 服务器市场竞争概貌

2001 年 Intel Itanium 架构第一代处理器 Itanium 正式上市前，市场上的服务器主要基于两类处理器：工业标准的 IA-32 处理器和 64 位 RISC 处理器。基于前者的服务器主要以 Windows 操作系统+Intel 芯片的 Wintel 标准占有低端入口级市场的主要份额，基于后者的服务器在高端企业级市场占有统治地位。虽然 64 位 RISC 服务器在低端也占有一席之地，但是很难与 Wintel 的性价比优势相抗衡；32 位服务器在性能、RAS 和可管理性以及技术服务等方面也与 64 位 RISC 服务器有很大差距，因此也很难进入最高端市场。其间的中档市场则是此消彼长的竞争空间。

Intel Itanium 处理器系列上市改变了服务器市场原有的格局，形成了三雄并列的竞争态势。其中，新兴的基于 Itanium2 处理器的服务器具有最强大的生命力。目前已经有 40 多家 IHV 厂商提供基于 IPF 的服务器产品，包括 HP、IBM、SGI、NEC、Unisys、Bull、Dell 等许多著名的厂商，使 IPF 成为事实上的工业标准架构。特别是，2003 年 HP 推出了基于 Itanium2 完整的服务器系列—Integrity 系列后，提供从 2 路入口级服务器到 64 路 Superdome 企业级服务器(使用 HP 专利的 mx2 模块可以扩展到 128 路)，使人们看到了工业标准服务器应用时代的曙光及其在性能和性价比等方面诱人的魅力。在 HP 等厂商努力下，短短 2 年中基于 IPF 服务器产品已经完成 64 位 RISC 服务器化了 5 年多完成的生态系统建设任务，包括开发出整套的系列产品、操作系统、编译程序和开发工具、数据库、系统基础件和业务基础件平台、面向各行各业的丰富解决方案(详见[14])。当前，基于 IPF 的新产品如雨后春笋出现在市场上，这一平台上应用软件和成功实例也与日俱增。虽然 IPF 服务器要全面夺取 64 位 RISC 的传统市场还需要假以时日，但是工业标准和开放性的 Intel Itanium 架构必将成为支持服务器(和高端工作站)应用的主流平台已经成为 IT 产业界和分析界公认未来发展趋势。根据 IDC 和

Gartner 等权威机构分析：在未来 3 年左右时间内，IPF 服务器将在中高端应用领域中占据主要的市场份额。随着 IPF 处理器批量的扩大和价格的下降，IPF 服务器在入口级也将得到长足的发展。

64 位 RISC 服务器厂商显然处于退守之势：HP、SGI、NEC、Unisys 等一大批厂商已经宣布在适当时间把服务器产品转向工业标准的 IPF 和 IA-32 平台。剩下的主要厂商只有 IBM 和 Sun 两家都处于两难的境地：IBM 一方面继续推出性能更高的 Power4+ 处理器，计划推出 Power5；另一方面，又推出基于 Itanium2 的新产品。IBM 目前正处于首鼠两端的摇摆境地，很可能会丧失良好的时机。Sun 也宣布要推出基于在 IA-32 基础上扩展的 X86-64 架构 Opteron 处理器的产品，已经开始在它的 64 位 RISC - UltraSparc 之外另觅出路。当然，厂商专属的 64 位 RISC 服务器市场不会突然消失，但是在工业标准的 IPF 和 IA-32 两边夹击下，必将一步一步走下坡路。

IA-32 服务器仍将在低端和桌面有其性能能够满足要求、价格低廉的优势，保持其传统地位。随着 Xeon 和 Xeon MP 等具有强大支持多处理器系统功能芯片的问世，基于这一架构的 HP ProLiant 等服务器系列也能够提供很高的性能。IA-32 服务器仍将与 64 位 RISC 服务器和 IPF 服务器在中低端剧烈竞争。随着网络游戏、流媒体、多媒体、数字特技等应用的普及，IA-32 的 32 位架构已经出现了明显的局限性。于是出现了 AMD 的 X86-64 架构，在保持 IA-32 基本结构和兼容性基础上，增加 64 位寻址功能的 X86-64 架构和 Opteron 处理器。X86-64 虽然有与 IA-32 兼容、64 位寻址和价格低廉等优势，但由于与 Itanium 或任何 64 位 RISC 处理器都不兼容、系统建设以及 IHV 和 ISV 支持以及 AMD 实力都与 Intel 的 IPF 处理器系列有很大的差距，其前景还很不明朗。目前只有很少的基于 Opteron 的服务器进入中低端市场。这种产品究竟将在 Intel 工业标准的 IPF 和 IA-32 夹击下走向消亡，还是能够在剧烈的市场竞争中占有一席之地，人们正拭目以待。

表 1-1 HP Integrity 服务器系列概貌

	入口级服务器			中档服务器		企业级服务器
服务器	rx2600	rx4640	rx5670	rx7620	rx8620	Superdome
处理器	Itanium2			Itanium2		Itanium2
主频	1.3 GHz , 1.5GHz			1.3 GHz , 1.5GHz		1.3 GHz, 1.5 GHz
个数	2	2-4	2-4	2-8	2-16	16,32,64
内存容量	1-24 GB	1-64 GB	1-96 GB	2-64 GB	2-128 GB	128, 256,512GB
系统带宽	8.5 GB /s	12.8 GB/s	12.8 GB/s			16, 32, 64 GB/s
I/O 插槽	1 个独立 PCI-X 插槽 ,3 个共享 PCI-X 插槽	2 个独立 PCI-X 插槽 , 2 个共享 PCI-X 插槽	3 个独立 PCI-X 插槽 , 7 个共享 PCI-X 插槽	15 插槽 (14 个双总线, 1 个单总线 PCI-X 插槽)	15 插槽 (14 个双总线, 2 个单总线 PCI-X 插槽)	48, 96, 192 个插槽(其中 2/3 为标准 PCI-X 插槽, 1/1 为高带宽 PCI-X 插槽, 例如 64 CPU 的 Superdome 有 128 个标准 64 个)

						高带宽 PCI-X 插槽)
I/O 带宽	2.5GB/s	3.0GB/s	6.5GB/s	15.4 GB/s	15.9 GB/s	8, 16, 32 GB/s
体系结构	SMP (对称多处理器)			cc-NUMA(缓存一致性 的非均匀内存访问)		cc-NUMA
互联机制	交叉交换		多层交叉交换		多层交叉交换	
芯片组	zx1		sx1000		sx1000	
扩展模块	mx2 , 使用 HP 专利的 mx2 模块可以在机箱内把服务器支持的处理器个数增加 1 倍					

表 1-2 当前市场上主要入门级服务器产品简表

厂商名称	RISC	Intel IA-32	Intel IA-64(IPF)
Bull	Escala S Series S120 Escala PL220R,T Escala PL240R,T Escala PL400R,T Escala PL420R,T	Express5800 TM600 Express5800 TM1400 Express5800 120Ef,Lg,Rd-1,Rf-1,Mf Express5800 140Hd,Rc-4 Express5800 320	NovaScale Model 4040
Dell		PowerEdge 400SC, 600SC, 650 PowerEdge 1600SC, 1655MC, 1750 PowerEdge 2600, 2650, 4600 PowerEdge 6600, 6500	PowerEdge 3250
Fujitsu Siemens	PRIMEPOWER 100N PRIMEPOWER 200, 200F PRIMEPOWER 250, 400PRIMEPOWER 400N, 450	PRIMERGY TX150, 200, 300, 600 PRIMERGY C200, L200, F250, H250 PRIMERGY RX 100, 200, 300, 600 PRIMERGY P200, 240, R450	PRIMERGY RXI600
Fujitsu	PRIMEPOWER 200, 250 PRIMEPOWER 400, 450	PRIMERGY TX150, 200, 300, 600 PRIMERGY C200, L200, F250, H250 PRIMERGY RX 100, 200, 300, 600 PRIMERGY P200, 240, R450, MS610	PRIMERGY N4000

HP	HP carrier-grade rp2450 HP server rp 2405, 2430, 2470 HP server rp 5405, 5430, 5470 AlphaServer DS10, 15, 20L, 25 AlphaServer ES45, 47	HP server tc2120 ProLiant BL10e, 10e G2, 20p G2, 40p ProLiant DL320, 320 G2, 360 G3 ProLiant DL560, 580 G2 ProLiant ML150, 310, 330 G3, 350 G3 ProLiant ML370, 530 G2, 570 G2	Integrity rx2600 Integrity rx4640 Integrity rx5700
IBM	eServer BladeCenter JS20 RS/6000 43P-150, 170 eServer pSeries 610, 615 eServer pSeries 630, 640 eServer iSeries 400 Model 250, 270 Model 800, 810, 820	eServer xSeries 205, 225, 235, 255 eServer xSeries 305, 335, 345, 360 eServer BladeCenter HS20	eServer xSeries 382 eServer xSeries 450
Maxdata			Platinum 9000R
NEC		Express5800 TM600 Express5800 TM1400 Express5800 120Ef,Lg,Rd-1,Rf-1,Mf Express5800 140Hd,Rc-4 Express5800 320La, 320Lb Express5800 420La, 410Ma	
SGI	Onyx 350, 3500 Onyx4 UltimateVision Origin 300, 350		
Stratus	Continuum 419, 429, 439, 449 Continuum 616, 616S Continuum 651-2, 652-2 Continuum 1251-2, 1252-2	ftServer 3300, 5600, 6600	
Sun	Fire B100 Blade Server Fire V100, V120, 280R Fire V210, V240, V250 Fire V440, 480 Netra 120 , Netra 20 Server	V60x Server, V65x Server	
Unisys		ES3005 Blade Server ES3020, 3020L, 3040,	

		3040L ClearPath HMP LX6100, LX7100	

表 1-3 当前市场上主要中档服务器产品简表

厂商名称	RISC	Intel IA-32	Intel IA-64(IPF)
Bull	Escala PL600T, 600R Escala PL800T, 820R, PL1600R	Express5800 180Rc-4 NoveScale9080	NovaScale5080, 5160
Dell		PowerEdge 8450	
Fujitsu Siemens	PRIMEPOWER 650, 800 PRIMEPOWER 850, 900	PRIMERGY RX800, T850	
Fujitsu	PRIMEPOWER 650, 800 PRIMEPOWER 850, 900	PRIMERGY RX800, T850, RX800 Trimetra P2000	
HP	AlphaServer ES80, GS80, GS160 Hp server rp7405, 7410, 8400	ProLiant DL740, 760 G2	Integrity rx7620 Integrity rx8620
IBM	eSrever pSeries 650, 655, 670 eSrever iSeries 400 Model 825, 830, 870	eSrever xSeries 445 eSrever xSeries 455	
NEC		Express5800/180 Rc-4	Express5800 1080Rc
SGI	Origin 3200		Altix 3300
Sun	Fire V880z Visualization Server Fire V880, V1280, 4800		
Unisys		ES700 Aries 510, 520 ClearPath HMP IX6600 ClearPath Plus Dorado Model 110	ES7000 Aries 410 ES7000 Aries 420 ES7000 Orion 430

表 1-4 当前市场上主要企业级服务器产品简表

厂商名称	RISC	Intel IA-64(IPF)
Bull	Escala EPC2450, PL3200R	NovaScale5080, 5160 NoveScale9080

Fujitsu Siemens	PRIMEPOWER 1000,1500,2000,2500	
Fujitsu	PRIMEPOWER 1000,1500,2000,2500	
HP	AlphaServer GS320, GS1280 AlphaServer SC20, SC45 HP9000 Superdome NonStop S74, S76SE, S740, S7400 NonStop S74000, S76000	Integrity Superdome
IBM	eServer pSeries 690 eServer iSeries Model 840, 890	
SGI	Onyx 3800, Origin 3800, 3900	
NEC		Express5800 1320xc Express5800 1320xd
Sun	Fire 6800, 12000, 15000	
Unisys		
注：如上所述，基于 IA-32 服务器很难进入企业级层次		

## 1.2 服务器市场需求和竞争势态分析

计算机问世的初期主要应用于科学计算,形成了完整的高性能技术计算(HPTC)应用领域。20世纪70年代以后,计算机应用逐步转向以信息处理为主,形成了规模庞大得多的企业应用领域。出现了一类新的计算机系统-服务器系统,在企业网络中提供后端的支持和服务。随着Internet的爆炸性发展和eBusiness技术的快速推广,各个档次的服务器逐步成为计算机应用领域的主角。当前,HPTC和企业应用在技术上互相渗透、计算量和数据规模越来越大、结构越来越复杂,对支持这些应用服务器系统,特别是高端的企业级服务器也提出了越来越高级和全面的需求。如何以更富有发展前景的技术生产出性能和性价比更高的产品全面满足用户需求、赢得更大的市场份额,已经成为各厂商竞争的热点;如何挑选最适合自己的服务器、特别是企业级服务器,也成为广大用户关心的焦点。其中最大的关键是需求分析。客户必须了解为了使得IT技术能够成为促进企业发展的能力,究竟需要服务器系统具有哪些特性以及它们的优先次序;厂商不仅必须了解行业的需求,而且必须了解客户的具体需求,生产出能够满足多样可变需求的灵活和可伸缩性产品。分析界也将根据市场需求来评估各种技术和产品。

### 1.2.1 高性能技术计算(HPTC )应用需求

高性能技术计算是人类探索和预测未知世界的工具,进行科技竞争的利器,也是国家综合国力的象征之一。HPTC技术广泛应用于核武器研究和核材料储存仿真、生物信息技术、医疗和新药研究、计算化学、GIS、CAE、长期气象和灾害预报、工业过程改进和环境保护等许多领域。过去,大型的HPTC应用主要使用专门制作的超级计算机。20世纪90年代

中期以后，随着 64 位 RISC、Internet/intranet 和分布式计算技术的发展，HPTC 应用转而大量使用基于 64 位 RISC 处理器的服务器。例如，最尖端的核技术和基因技术研究都大量使用基于 Alpha 的服务器。当前基于 Itanium2 的服务器已经在这一领域取得了巨大进展，特别是利用低端 Itanium2 服务器或工作站作为节点的集群系统已经在 HPTC 领域占有重要地位。HPTC 应用对服务器系统的需求主要是：

- **计算能力强**：能够在最短的时间内完成最大的计算量，特别是高精度(字长 64 位、128 位或更高)浮点计算；
- **储存容量大**：能够在内存和磁盘中储存最大数量的信息，特别是提供处理容量高达几十以至几百 GB 的内存数组能力的超大规模内存(VLM)技术，对 HPTC 应用具有很大的实用价值；
- **系统带宽高**：能够以最快的速度在处理器和内存、内存和磁盘之间传输信息。由于处理器速度远远高于内存系统，系统是否能够提供足够的带宽，保证内存能够及时向多个处理器提供足够的数据，使得处理器觉察不到内存系统的延迟(即没有可觉察延迟)，是提高系统性能的关键；

### 1.2.2 企业应用需求

企业应用一般也把计算机在政府部门、国防军事、研究院所等领域的管理应用包括在内，是以信息处理为主应用的统称。此类应用系统早期称为管理信息系统(MIS)，以后演变为生产资源规划(MRP-II)、企业资源规划(ERP)、计算机集成制造系统(CMIS)、产品全过程管理系统(PLM)和决策支持系统(DSS)等等。随着 Internet 的发展和人类步入信息时代，企业应用正向电子商务企业(eBusiness)方向发展。企业应用的模式可以用分布式数据库应用来概括，从早期的利用一个主机支持通过广域网(WAN)联接的大量终端的分时主机模式、到利用服务器或服务器集群支持通过 WAN/LAN 联接的大量客户机的客户机/服务器(C/S)模式、发展到目前的利用整合的服务器或集群系统支持通过 Internet/intranet/extranet 联接大量 Web 浏览器的 eBusiness 模式或 3 层模式(或多层模式)，表面上都是利用某种服务器系统支持网上大量分布式数据库用户，实质上面向大型企业(如大型企业集团、跨国公司、政府部门)的高端企业应用的复杂性和综合性以及对服务器要求已经发生了本质的变化。

早期的企业应用与 HPTC 应用有很大的差别，使用的数据量很小、计算也比较简单，对计算机系统的主要要求是能够支持大量用户进行事务处理如信息输入、查询和统计等。系统能够支持同时上网的用户数量越多越好、延迟时间越短越好，至于计算能力、存储容量、系统带宽往往要求不高、或者只是某种间接的要求。这与对它们需求极其迫切的 HPTC 应用截然不同。因此，不同类型的用户在设备选型时路线也大不相同。例如，很难想象企业应用用户会从 Cary 或 Convex 公司、HPTC 用户会从 Wang 公司购买计算机。

当前，人类正在从工业化社会进入知识经济的时代，技术的持续创新、市场需求的瞬息万变、竞争空间的迅速扩大，要求企业采用 Internet、电子商务、电子商务企业等现代化手段，来适应时代的发展。企业应用对计算能力、存储容量和系统带宽的需求也变得越来越高、越来越迫切，企业应用与 HPTC 应用的界限也越来越淡化。这些变化主要表现在如下方面：

- **从运行单一应用到运行复杂的应用**：过去不少企业信息系统往往运行单一的应用，但支持 eBusiness 的企业系统要求服务器支持企业电子商务基础设施(eInfrastructure)，运行十分复杂的应用，包括、电子商务、客户关系管理(CRM)、商务智能(BI)、全球价值链(GVC)、知识管理(KM)、协作和信息传递等方面的应用。这些应用组合在一起，构成了十分复杂的应用环境；

- **从面向本地到面向全球竞争的一体化应用：**为了在全球竞争中赢得生存和发展的机会，新型的电子商务企业把整个业务流程看作是一个全社会甚至全球范围内紧密联接的价值链(或供应链)。它不仅要求利用 Internet/intranet 联接和管理本企业整个供应链上所有环节，而且还要求覆盖供应商、制造工厂、合作伙伴、分销网络和客户，甚至还包括对竞争对手的监视和分析。供应链上的环节包括订单、采购、库存、计划、生产制造、质量控制、运输、分销、服务和维护、财务管理、人事管理、实验室管理、项目管理等等。全球价值链技术的应用大大扩展了企业系统覆盖的范围，提高了企业系统规模和信息流量；
- **从单一媒体发展到多媒体：**当前的企业应用已经由单一的字符媒体发展到处理文字、图形、图象、声音、视频、流媒体等多媒体信息；
- **从手工输入发展到自动输入：**当前信息输入的方式已经从主要通过手工方式发展到把自动数据采集，人们通过扫描、摄像、遥测、遥感等数字化手段，把大量信息输入到系统中进行处理；传感器和传感器网络已经成为许多应用最重要的数据来源；\* 从 OLTP 发展到 OLAP：过去企业系统的应用模式主要是基于数据库从在线事务处理(OLTP)发展到基于数据仓库的在线分析处理，通过数据采掘、商务智能(BI)等手段，帮助企业更加充分地利用现有的信息如发现最有前途的市场领域等；
- **从信息查询发展到辅助决策支持：**现代的企业信息系统不仅支持信息查询而且通过提供全面信息和计算工具，支持企业领导作出正确的决策；
- **从一般的可用性发展到支持连续运行的高可用性：**电子商务的发展使得许多企业信息系统每停机 1 小时都会造成巨额的损失，因此要求 99.999% 的高可用性，以及与智能化自动管理、容错和容灾功能相结合的支持连续运行的高可用性，以满足大量关键任务应用的需求；
- **从允许延迟发展到“零延迟”：**电子商务时代任何延迟往往会造成丢失客户、降低效率或者遭受欺诈损失，因此要求企业系统尽可能降低延迟，向“零延迟”(ZLE)方向发展；这些变化反映了信息技术应用的未来发展趋势，也从根本上改变了企业应用对服务器系统的需求。对服务器在计算能力、内存和磁盘容量、内存和 I/O 带宽等方面都提出全面的高要求。如果说在过去以衡量 OLTP 吞吐能力(分布式数据库应用环境中对大量同时用户的响应能力)的 TPC-C 为主要基准测试指标来衡量企业应用服务器还有一定的道理的话，这种观点现在已经不适用了，也不符合未来的发展趋势。既然企业应用对服务器的需求是全面的，那么衡量服务器的性能指标也必须是全面的。

### 1.2.3 应用需求的融合

当前，除了极少数最高端的 HPTC 应用(此类应用由于种种原因，往往不惜一切代价来追求高性能，抢时间、争速度) 以及规模较小的低端企业应用外，企业应用和 HPTC 应用对服务器系统的需求正在逐步融合在一起。许多 HPTC 系统也要求具有较强支持大量分时用户的能力。例如，支持基因研究的全球性协作的系统、支持工厂生产流程自动化的 CAE 应用等。企业应用系统在要求具有较高的分时应用吞吐能力和响应速度外，也变得象 HPTC 系统那样同样需要很高的计算能力、存储容量和系统带宽。不少信息处理应用如地理信息系统(GIS)、化学信息学、生物信息学等已经归入 HPTC 应用的范畴，甚至成为最高端的 HPTC 应用如人类基因密码破译等。此外，还出现了把支持产品设计生产的 CAD/CAM 应用与企业全球价值链结合在一起的产品全过程管理(PLM)等新型应用。随着两者需求的融合，服务器产品的一个很明显的发展趋势是向通用化方向发展。事实上，目前所有主要厂商都生产通用的服务器产品，面向广阔的市场领域。一些以面向专门市场领域产品著称的厂商都已经或者退出市场或者失去了昔日的光辉。

#### 1.2.4 对服务器系统全面的需求

当前人类社会已经从工业化时代进入知识经济和信息化时代，随着经济的全球化，技术不断的创新、市场需求瞬息万变、企业竞争空间迅速扩大和全球化，程度日益剧烈、节奏也不断加快。许多企业都不得不越来越依赖于基于 Internet 和 eBusiness 模式的信息技术来求得生存和发展，支持企业信息系统的服务器及其生产厂商也提出了越来越高和全面的要求。这些要求可以归结为性能指标、系统特性和厂商能力等三大类，具体如表 1-5 所示。

上述各项是针对经济全球化和信息化条件下以企业生存和发展的实际需求为背景提出的，在分析服务器技术和服务器选型时必须予以全面综合考虑：

- **全面的性能指标：**如上所述，现代的企业应用系统类型广泛、规模大、综合性强，要求服务器系统具有全面的高性能：速度更快、内存和磁盘容量更大、系统带宽和 I/O 及网络联接能力更高。因此，必须使用全面的指标来衡量和比较系统的性能。任何一项基准测试指标都有其局限性和片面性。有的厂商往往按照自己的优势领域来诠释，过份强调某项指标、贬低或不公布其他指标，这样很容易误导用户。事实上，性能指标都是以系统资源容量(处理器、内存、磁盘、I/O 接口等)以及系统架构是否能够充分发挥资源潜力为基础的。只有从实际(系统资源容量和架构)出发，全面考察系统的基准测试指标，才能作出科学和客观的评估，选择最佳的服务器；
- **先进的系统特性：**企业应用对系统特性的要求也反映了全球经济和电子商务时代特色。企业要在日趋剧烈的竞争中生存和发展，必须要求所使用的服务器具有先进的系统特性。例如，高可伸缩性反映了企业满足业务发展需要和保护原有投资的要求；高可用性反映了企业对支持电子商务应用系统 7\*24 连续运行的要求；高可管理性反映了企业提高设备利用效率和加速投资回报的要求等等。当前，尽管所有厂商都声称其厂商具有最先进的系统特性。但事实上，由于技术水平和发展历史的差异，不同厂商产品的可伸缩性、RAS 和可管理性水平是有很大差别的。因此，在分析服务器和进行服务器选型过程中必须结合业务需求、既全面又区分轻重缓急地考察服务器产品的系统特性；
- **厂商战略、实力和水平：**现代企业信息系统是一个技术日新月异、规模迅速扩大、应用逐步深入的发展中系统，也是企业正常经营不可缺少的基础条件。因此，企业信息系统建设成功经验之一是把关键设备如服务器。特别是企业级服务器供应商作为信息技术方面的长期合作伙伴。因此，在选择企业级服务器的同时，必须全面考察厂商产品战略、经济实力和技术水平，包括产品发展战略是否与企业的长期目标一致、厂商提供全面解决方案能力、应用软件数量和质量以及系统平台上 ISV 和合作伙伴队伍规模、在本行业实际应用经验和成功实例、提供技术支持和服务的能力等。对为企业战略目标服务的企业级服务器，对厂商本身的全面考察甚至比产品具体指标的某些差异还要重要；

### 1.3 HP Integrity 系列服务器架构及其实施

HP 基于 Itanium2 的 Integrity 系列服务器是针对当前企业用户对服务器的需求而优化设计的。HP 的产品总体战略是在工业标准技术基础上进行创新和增值。这一战略奠定了 HP Integrity 系列的设计思想的基础。一方面，HP 是工业标准处理器平台最积极和有力的支持者，制订了明确的战略和发展蓝图、把高端服务器产品全面转向 Intel Itanium 体系结构的 IPF 处理器系列。HP 使用工业标准 Itanium2 处理器、模块化设计和标准化部件为服务器产品提供领先性能和性价比奠定了坚实的物质基础；另一方面，HP 对基于工业标准的服务器产品进行了一系列创新和增值，包括采用适合系统资源规模和应用的体系结构、互联拓扑，基于 HP 独特的芯片组最能够发挥硬件潜力的软件进行系统实施，为用户提供最广泛的操作系统选择空

间，开发丰富的系统和应用解决方案等，从而确立了 HP 在工业标准平台上的技术优势，奠定了 HP Integrity 系列各个档次服务器在性能、性价比和企业应用特性等方面的竞争优势和市场领先地位。

### 1.3.1 HP Integrity 服务器系列的设计思想

为了满足应用的需求，必须通过把多个处理器模块、内存模块和 I/O 控制模块互相联接构成庞大和复杂的计算机系统。为了更加有效地发挥所有组成部件的潜力，组成高性能、高可伸缩、高可用的多处理器系统，必须事先有一个考虑得很周密的设计蓝图，来指导如何把各种类型的计算机资源以及系统软件集成在一起。这就是计算机架构要解决的问题，即按照怎样的模式把各个部件联接成一个完整的计算机系统。在既定的架构指导下，还必须决定系统内部的互联拓扑并且利用相应的芯片组来实施系统的架构和互联拓扑。针对厂商普遍采用工业标准处理器和模块化构件来设计和生产服务器的发展趋势，HP 制订了在工业标准基础上进行增值的产品发展战略：采用工业标准使产品具有开放性和批量优势，进行增值使产品具有独特的优势，两者相辅相成、缺一不可。HP Integrity 服务器系列基于工业标准的 Itanium 架构的 IPF 处理器，与此同时 HP 又利用其先进的系统技术和丰富经验，对系统硬软件进行了全面的增值。其中最核心和基础性的是在架构、互联拓扑以及实施芯片组方面的精心设计和增值，奠定了 Integrity 系列提供一系列独特优势的技术基础。

“架构”一词是从英语 “architecture”一词翻译过来的，其原意是“建筑学”或“建筑设计”。事实上，按照某种架构建造一个计算机系统，就好象建筑行业中利用各种建筑材料，建造一座大厦。HP 采用符合当前计算机技术发展潮流的设计思想，为 Integrity 产品系列各个档次系统选择了最适合应用和市场需求的最佳架构：Integrity 入口级服务器采用 UMA 架构，简化了硬软件的设计、降低了系统开销，实现在确保高性能条件下，提供最佳的价格/性能；中高档服务器采用 ccNUMA 架构，允许用户按照 UMA 相同方式编程，同时提供最大的可伸缩性。

当前许多应用领域都要求服务器系统具有多处理器、大规模内存和高 I/O 吞吐能力以及高可伸缩性，处理器之间以及处理器与内存模块和 I/O 控制器之间的通信要求和难度也越来越高。事实上，所有服务器系统都通过互联设备扩大其规模，形成丰富的系统资源。为此，服务器系统必须根据其架构、设计高效、经济和先进互联拓扑(即处理器、内存模块和 I/O 控制器之间的联接关系和通信路径)，方能够更加高效地实现并行处理，把系统资源的潜力变成高性能指标和真正的实际信息处理能力，同时提供最高的可伸缩性和可用性以及最佳价格/性能。因此，多处理器互联技术已经成为建立可伸缩并行系统的最关键的技术、服务器系统成败优劣的决定性因素之一。HP Integrity 系列的互联拓扑采用了先进性与可行性相结合的原则，在入口级和中高档分别采用交叉交换和分两层交叉交换的互联拓扑，实现了既确保高性能又提供高性价比的设计思想。

芯片组是实施服务器架构和互联拓扑的关键器件。随着芯片组在服务器系统中的枢纽作用的凸现，芯片组技术包含的服务器核心逻辑正在成为服务器的核心技术。许多芯片厂商和系统厂商都在研发自己的芯片组产品，使得基于 IA 服务器差异化增强，呈现百花齐放的局面。HP 在 HP 9000 系列服务器成功经验的基础上，分别使用 zx1 和 sx1000 芯片组实施 Integrity 入口级和中高档服务器的总体设计，并且以创新的 mx2 模块倍增服务器所支持的处理器数量。

HP Integrity 系列领先的架构、互联拓扑和芯片组技术，奠定了以工业标准部件构建具有最高性能和性价比计算机产品的基础，形成了领先的全系列产品，开创了 64 位工业标准计算的新时代。

### 1.3.2 HP Integrity 入口级服务器的设计和实施

HP Integrity 入口级服务器采用适合于低端系统资源规模、价位和应用特点的 UMA 架构和交叉交换互联拓扑，并且以领先的 zx1 芯片组实施这一总体设计，从而赋予 Integrity 入口级服务器一系列竞争优势。

## UMA 架构(均匀内存访问架构)

UMA 架构也称为 SMP 架构(对称多处理器架构)，是一种使用共享和集中内存、所有处理器和内存模块通过适当的互联设备对称地联接在一起的架构。Integrity 入口级服务器采用 UMA 架构既能够满足 4 处理器以下低端系统的扩展需要(不会由于 UMA 系统可伸缩性较差，而影响系统的扩展特性)，又便于实施、避免处理本地和远程带来的复杂性和开支，确保高性价比和入口级层次的高性能。

## 交叉交换互联拓扑

传统的 UMA 系统大多数采用总线互联拓扑。其优点是比较容易实施，其缺点是所有 CPU 共享相同的与内存的通信路径。随着 CPU 数量的增加、速度的提高，一些 CPU 不得不“停下来”等待内存读写，这将大大降低系统的吞吐能力和性能指标。特别是对于 Itanium2 这样的高性能处理器即使在入口级采用总线互联拓扑也会妨碍 CPU 充分发挥潜力。因此，Integrity 系列从入口级开始即采用交叉交换互联拓扑，在 CPU 之间或者 CPU 与内存之间建立多个并行的双向通信路径。这些路径是点到点的(非共享的)，数据能够在 CPU 和内存之间连续传输，没有竞争和等待问题、系统延迟时间也不会随着工作负载的增加而增加。

Integrity 互联拓扑的设计思想是把入口级与中高档分开：入口级采用一层交叉交换互联拓扑，中高档采用两层交叉交换互联拓扑。这样做在入口级既防止了总线拓扑的资源竞争、又避免了建造大交叉交换器的开支，实现了性能和性价比的最佳结合。

## zx1 芯片组

HP 使用 zx1 芯片组实施 Integrity 系列入口级服务器的 UMA 架构和交叉交换互联 拓扑。该芯片组基于清晰的模块化设计，由核心部件(zx1 内存和 I/O 控制器)以及两种选购部件(zx1 可伸缩内存扩展芯片和 I/O 适配器芯片)组成。核心部件用以满足规模较小 IPF 系统普遍需要，选购件提供满足高度优化专门设计需要的灵活性。

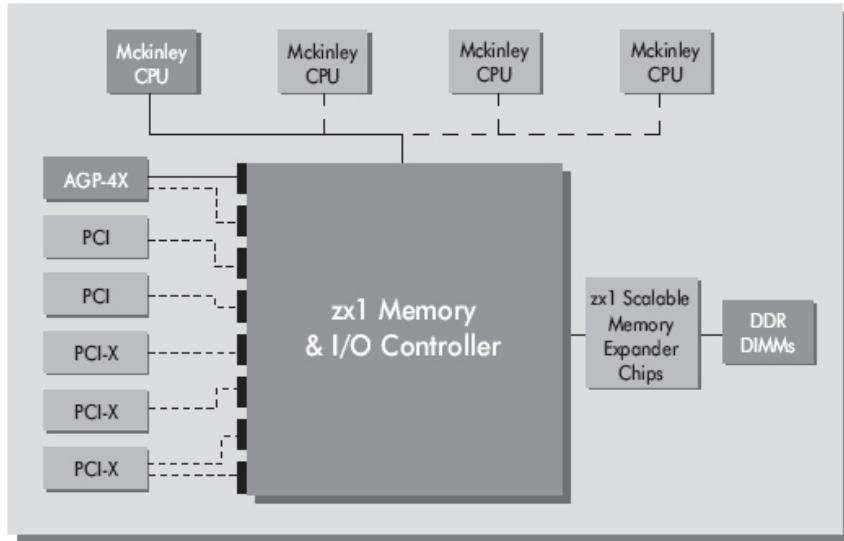


图 1-2 zx1 芯片组方框图

基于 zx1 计算机系统的内存可以直接或通过 zx1 可伸缩内存扩展芯片与 Itanium2 处理器总线相联接，提供处理器和 DDR 内存之间的低延迟联线。zx1 可伸缩内存扩展芯片可以选择地设计成同时增加内存容量和内存错误校正能力。它可以作为内存集线器，降低内存总线上信号负载、提高内存传输速率。

zx1 芯片组可以包含 1-8 个 I/O 适配器。每个 I/O 适配器支持一个 PCI, PCI-X 或 AGP-4X 总线。每个 I/O 适配器可以联接到来自芯片组核心的 1 条或 2 条 500 MB/sec I/O 联线。这一设计提供很大的灵活性：

- 当需要支持最大个数低带宽 I/O 设备时，可以把每个 I/O 适配器与一条联线相联接，在每条总线上装上最大个数的 I/O 插座；
- 当需要较好的故障隔离时，可以限制每个 I/O 适配器只有一个 I/O 插座；
- 当需要最大带宽时，可以在每个 I/O 适配器上接两条联线；

一个系统上可以使用任何组合满足不同的需求。这一选择灵活性允许系统设计人员利用同一芯片组设计满足各种实际需要的系统。最简单的配置只需要 zx1 内存和 I/O 控制器联接一个处理器、内存模块和 I/O 适配器；最大的配置可以联接 4 个处理器、12 个内存扩展器和 8 个 I/O 适配器，提供 64 个 DIMM 内存容量和 4 GB I/O 带宽。事实上，HP Integrity 系列入口级的 rx2600, rx4640 和 rx5670 服务器以及 zx2000 和 zx6000 工作站都是围绕 zx1 芯片组实施的。HP 的设计思路是使用专门设计的芯片组来满足实施低端和中高档系统的需要：使用价廉物美的 zx1 芯片组来实施低端产品，使用可扩展性更强的 sx1000 来实施中高档服务器。这样做好处是：

- 面向低端的 zx1 芯片组不必考虑 4 处理器以上的系统内可伸缩性和维持 NUMA 系统缓存一致性的功能，只需要提供实现入口级规模下的高带宽、低延迟和灵活性，确保 Integrity 入口级服务器的高性能和高性价比；
- 面向中高档的 sx1000 芯片组则必须实现以最低系统开销以及最低的访问本地合远程内存延迟时间比支持 cc-NUMA 架构，提供建立中高档系统的可伸缩性、高聚合带宽和灵

- 活性，
- 确保 Integrity 中高档服务器的高性能和高性价比；
  - 与其他一些厂商试图利用同一芯片组支持各个档次服务器的设计思路相比较，HP 的设计策略在经济和技术上都有其明显的，为 HP 率先实现提供从入口级到 64 处理器企业级 Integrity 服务器系列以及领先的性能和性价比奠定了坚实的基础；
  - 有人说，HP 的两个芯片组的策略不利于入口级服务器向上扩展。事实上，几乎没有用户会考虑把入口级服务器垂直升级到企业级，HP 为入口级服务器和工作站提供了极好的水平扩展的途径。HP Integrity 入口级服务器允许使用具有两条联线的 I/O 适配器支持高速的集群联接。目前，已经有许多用户利用 HP 基于 IPF 的低端产品建立 Linux 集群架构的超级计算机系统。其中，HP 与美国能源部西北太平洋国立实验室建立的 IPF-Linux 集群已经进入了 TOP10 的行列；

### 1.3.3 Integrity 中高档服务器的设计和实施

HP Integrity 中高档服务器包括中档的 rx7620 和 rx8620 服务器和企业级的 Superdome 服务器，它们采用适合于中高档系统资源规模、价位和应用特点的 NUMA 架构和两层交叉交换互联拓扑，并且以领先的 sx1000 芯片组实施这一总体设计，从而赋予 Integrity 中高档服务器一系列竞争优势。

#### NUMA 架构 (非均匀内存访问架构)

NUMA 架构是将若干个单元(有的系统也称为模块、构件、节点等)通过专门的互联设备联接在一起组成的分布式和共享内存架构。HP Integrity 系列中高档服务器采用高效的 cc-NUMA 架构：

- 使用专门的硬件把所有各个单元内存合并成一个统一的地址空间，各个处理器都可以通过通常的 Load/Store 指令直接访问整个地址空间，允许用户使用类似于 UMA 模式编程，也可以直接运行 UMA 架构下开发的应用软件；
- 每个单元都拥有自己的处理器和内存。因此，在 NUMA 架构下处理器和内存都是分布式的，从而支持整个系统通过增加单元、增加处理器个数、内存容量(以及 I/O 联接能力和带宽)，实现最大的可伸缩性；
- 采用高效的目录机制，来解决 NUMA 系统中的内存和缓存一致性问题。所以把 Integrity 中高档服务器的架构称为 cc-NUAM (缓存一致性非均匀内存访问) 架构；

总之，Integrity 中高档服务器通过采用 cc-NUAM 架构既保持 UMA 便于编程的优势，又提供中高档服务器所必须具备的可伸缩性，同时还能够高效地实现缓存一致性，把开销降到最低。

#### 两层交叉交换互联拓扑

随着处理器速度的提高，利用传统的总线拓扑来联接 8-16 个处理器，必然造成很大的系统开销和延迟。实践证明，总线拓扑很难扩展到支持 16 个以上的高性能处理器。同时，

利用一个大规模的交叉交换器来支持中高档服务器大量处理器和内存模块(以及 I/O),在经济上也是极不现实的。因此,HP Integrity 系列中高档服务器采用两层交叉交换互联拓扑。这样做既避免了构建大规模交叉交换器的开支和总线拓扑的开销,又能够以较低的成本实现高可伸缩性、高组合带宽、低系统延迟、消除系统瓶颈,实现中高档系统高性能和高性价比的需要(示意图见图 1-2)。

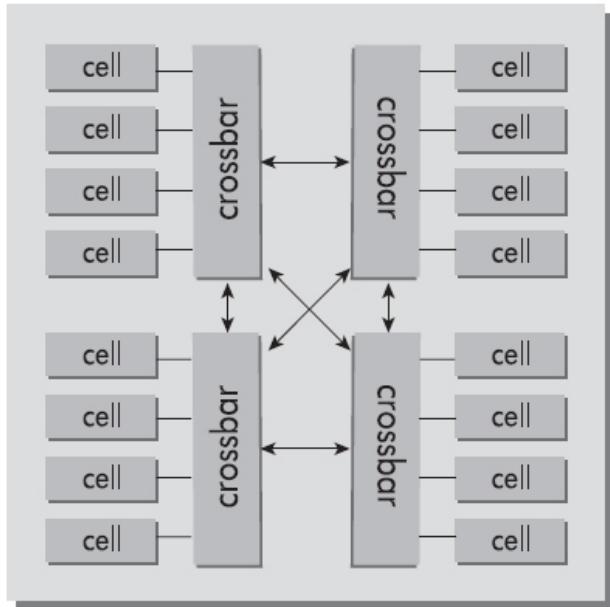


图 1-2 两层交叉交换互联

## sx1000 芯片组

HP Integrity 中高档服务器使用 HP 专门设计的 sx1000 芯片组来实施其 cc-NUMA 架构和两层交叉交换互联拓扑。

HP Integrity 使用专门设计的 sx1000 芯片组实施中高档服务器的 cc-NUMA 架构和两层交叉交换互联拓扑:

- 支持单处理器和双处理器两种类型的 CPU 模块: 使用单处理器 CPU 模块可以建立 2 到 64 路
- 高可伸缩的系统: 使用双处理器 CPU 模块 mx2 可以建立 128 路系统, 实现当前最高的扩展上限;
- 支持 Itanium 和 PA-RISC 两种类型的处理器: 具体地说, sx1000 芯片组将支持 Itanium 2 处理器(代码名字 Madison) 或 PA-RISC 8800 处理器;
- 支持灵活的 OS 环境: HP-UX, Linux, Windows Server 2003 和 OpenVMS;
- 提供投资保护: 为当前使用基于 PA-RISC 的 rp7400, rp8400 和 Superdome 用户提供平滑升级的路径, 这些系统现有内存、系统背板和 PCI 卡的单元和 I/O 子系统都可以继续使用, 以保护原有的硬件投资;

sx1000 是在支持 HP 9000 的 Yosemite 基础上发展起来的。在它的 5 种 ASIC 中, 单元控制器、内存缓存、PCI-X 系统总线适配器和 PCI-X 主机桥等四种提供更高性能和更新功

能；最后一种 ASIC 交叉交换器的性能和功能保持不变。表 1-7 列出 sx1000 5 种 ASIC 的特性和功能。

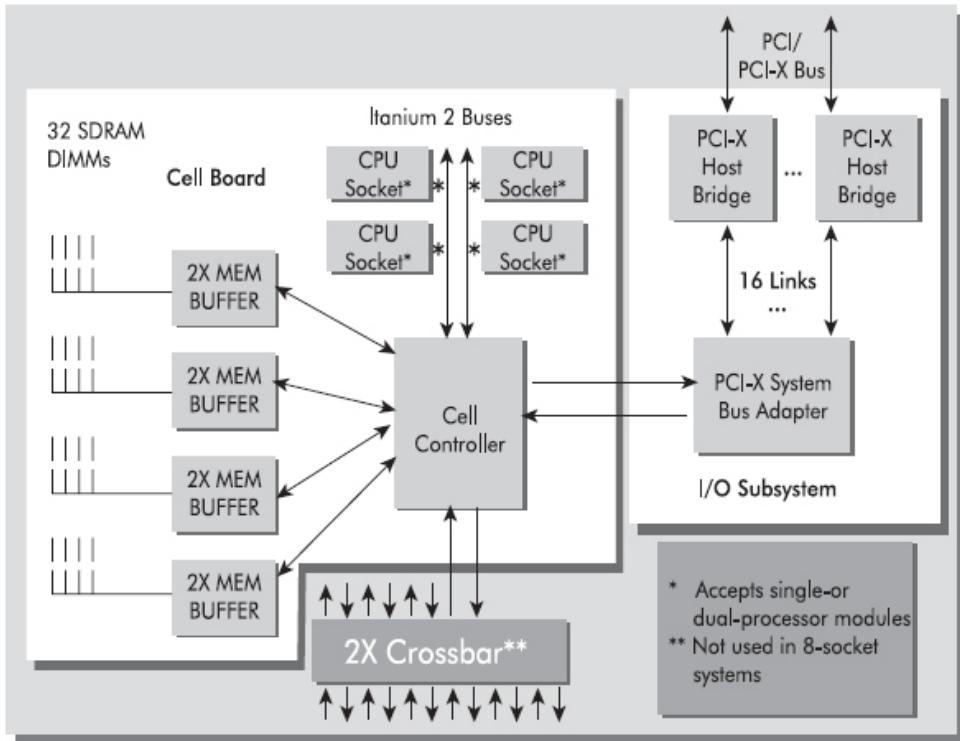
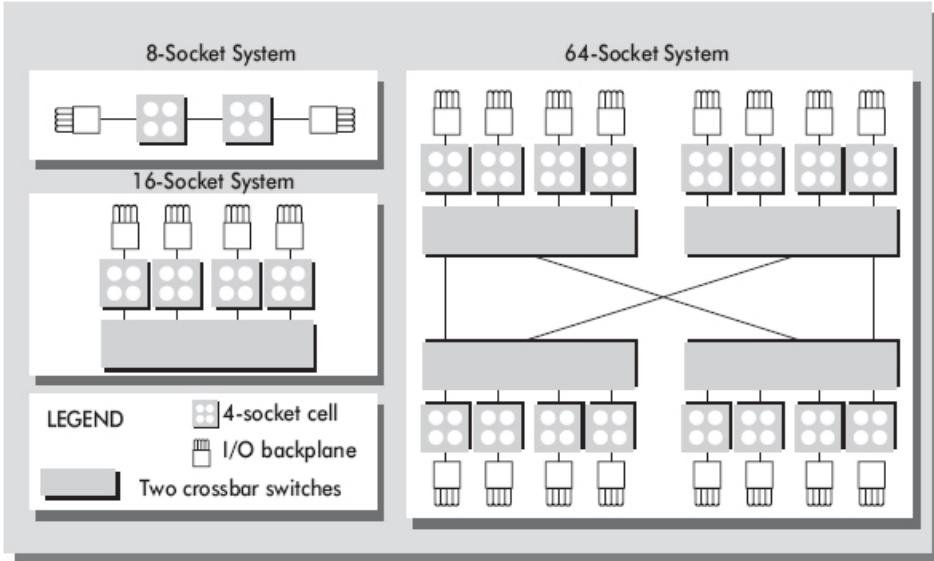


图 1-3 HP Integrity 单元逻辑方框图

Integrity 中高档服务器的基本构件是单元(cell)，单元建立在 sx1000 的单元控制器、内存缓冲区、等 ASIC 上，再利用 sx1000 的交叉交换器联接在一起构成 2-128 路高度可伸缩的服务器系统。每个单元由一块支持 4 个 CPU 插槽和 9 个 ASIC(1 个单元控制器和 8 个内存缓冲区芯片)的印制板组成。每个 CPU 插槽中可以插入一个包含 1-2 个处理器的 CPU 模块。使用 sx1000 芯片组的最小系统支持两个单元(8 个插槽)，最大的系统支持 16 个单元(64 个插槽)。

图 1-3 表示单元的方框图。其中，每个单元控制器负责单元的 4 个 CPU 模块之间的数据和信号传递，提供 4 个独立的内存子系统的控制和缓存管理功能。单元控制器通过交叉交换器联接到其他单元。单元控制器也联接到一个 I/O 后面板，PCI-X 系统总线适配器和 PCI-X 主机桥联接到 PCI 和 PCI-X 卡。PCI-X 总线适配器可以通过印制电路或电缆联接到单元控制器。

单元和 ASIC 交叉交换器装配成图 1-4 所示的 8 插槽、16 插槽和 64 插槽系统。交叉交换器成对使用，一对交叉交换器可以联接 4 个单元(16 个插槽)。16 插槽系统需要使用 1 对交叉交换器，64 插槽系统需要使用 4 对交叉交换器，而 8 插槽系统不需要使用交叉交换器，可以使用电缆直接联接两个单元。



**图 1-4 HP Integrity 系列中高档服务器组成示意图**

与业界其他芯片组相比较，sx1000 芯片组在技术上是相当先进的。所有 5 个 ASIC 芯片都是使用 0.18-micron CMOS 技术制造的。如表 1-8 所示，单元控制器包含 41 M 个晶体管，在 2500 针封装中提供 1466 个信号针。

### HP Integrity 系列服务器的扩展——mx2 模块

为了以最小的代价扩展 Integrity 系列各个档次服务器的规模，HP 提供专门设计的 mx2 双处理器模块实现：

- **更高的可伸缩性：**把服务器中 CPU 数量增加 1 倍；
- **更高的性能：**应用中实现 50% 以上性能提高，以最小的扩展成本满足 ERP、数据库应用和事务处理等企业应用的需要；
- **灵活性：**使用与标准 Itanium2 处理器模块相同的插座和电源；

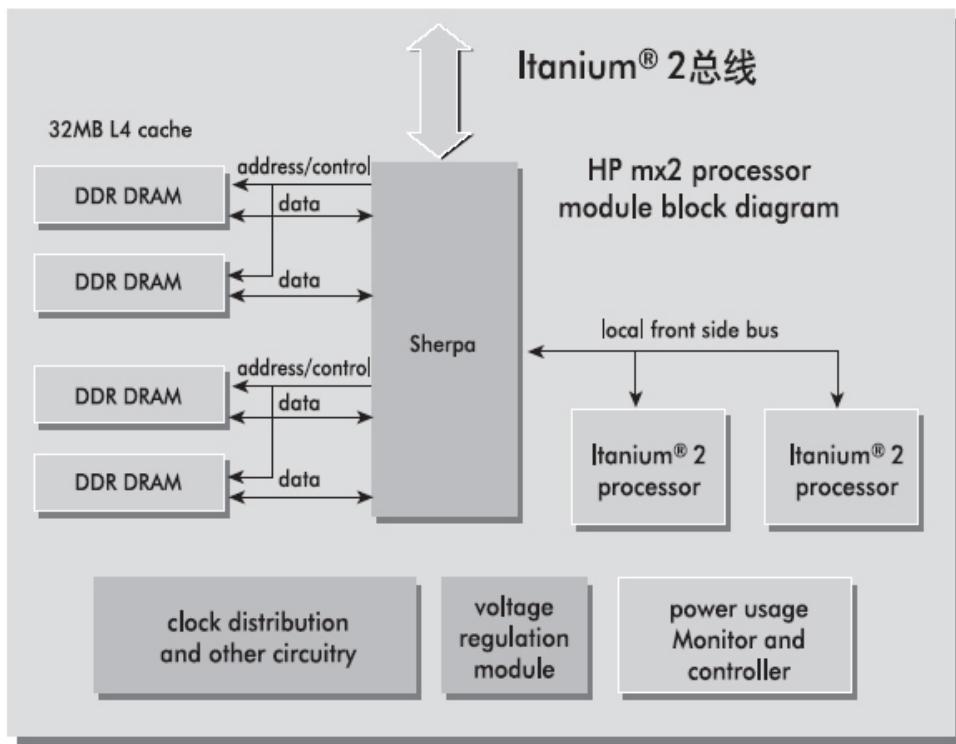


图 1-5 mx2 方框图

每个 mx2 模块包含 2 个 Itanium2 处理器和 32 MB L4 缓存。其内部布局如图 1-5 所示。

## 1.4 HP Integrity 系列的优势

HP Integrity 服务器系列体现了当前基于工业标准 IPF 处理器系统的最高水平，具有全面的技术和市场竞争优势，代表了当代支持高端应用的服务器产品未来发展方向。HP Integrity 服务器系列的优势可以分为两类：

- **Integrity 系列的共同优势：**HP 是 EPIC 技术发明者、IPF 系列的共同开发者。HP 坚信工业标准的 Itanium 架构代表 64 位处理器的未来方向，大力支持 IPF 处理器系列发展成为支持高端应用的主流处理器产品。HP 支持 IPF 的全面战略，技术上先行优势和丰富积累，在系统设计、应用开发和服务支持等方面的实力，使得 HP 的 Integrity 系列具有一系列共同的优势，如覆盖各种类型和各个档次应用的产品系列、高性能和性价比、提供多操作系统、高可伸缩性、高可用性和可管理性、丰富的解决方案和广泛的应用以及强大的技术服务和支持等；
- **各个档次产品的具体优势：**HP Integrity 系列是业界最完整的基于 IPF 产品系列，各个档次的产品由于系统资源规模、价位和应用领域的差异，有许多不同的特点。HP Integrity 系列不仅有许多整体优势，而且在各个档次上都具有一系列领先于其他厂商的具体优势。例如，具有适合于相应档次资源规模的架构，提供超过同档次竞争产品的性能和性价比，提供符合相应档次应用特点的领先特性如在入门级具有最高的性能密度、在企业级提供强大的整合功能等；

为了避免重复，本章将概要地介绍 HP Integrity 系列的共同优势，各个档次的具体优势将在以后的章节中介绍。

### 1.4.1 HP 支持 IPF 的全面战略

HP 已经作出了坚定的战略决策：在通盘考虑和兼容的 RISC 处理器产品的同时、把高端系统产品全部转向 IPF 平台。HP 的 PA-RISC 处理器系列在目前的 PA-8700 基础上再发展二代即 PA-8800 和 PA-8900 后、Alpha 处理器系列在目前的 EV68 基础上再推出 EV7、EX7z 和 EV79 等新产品后，都将不再继续发展，基于它们的 HPe3000 以及 HP-9000 和 AlphaServer 等服务器产品系列都将过渡到 IPF 平台上。HP 基于 MIPS 处理器的 NonStop Himalaya 服务器系列也将过渡到 IPF 平台上。至此 HP 的服务器产品将全部过渡到 Intel IA-32 和 Itanium 架构下。HP 不仅制订了明确的向工业标准架构过渡的发展蓝图、而且提供最平滑途径实施这一过渡，使 ISV 和用户的硬软件投资得到最安全的保护。与一些厂商举棋不定或者对未来发展趋势视而不见的战略相比较，HP 的战略代表了未来发展方向、是业界最坚定、最明确、最大规模的过渡到 IPF 发展蓝图，树立了 IPF 作为未来主流平台的旗帜，在计算机和信息产业界产生了巨大影响，吸引和鼓舞广大用户、ISV 和 OEM 厂商聚集在 IPF 的旗帜下。HP 的坚定战略、技术积累和雄厚实力奠定了在工业标准产品领域的领先地位坚实基础。

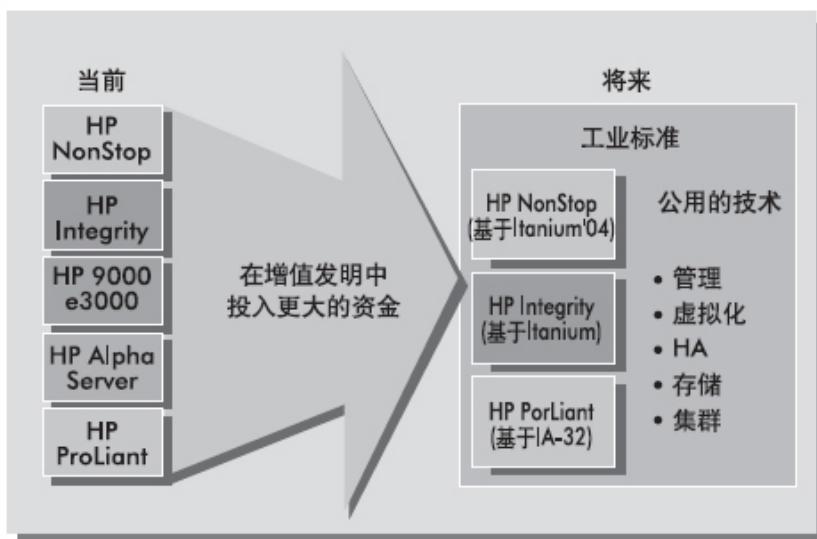


图 1-6 过渡到 3 个领先的产品线——建立在 2 个工业标准架构上

### 1.4.2 HP Integrity 系列完整的产品覆盖面

HP 向工业标准过渡的全面战略使 HP 集中力量开发基于 IPF 的产品。HP 率先推出了基于 Itanium2 的从入口级到企业级全系列产品，与 64 位展开全方位的竞争，充分显示了 IPF 处理器系列的优势和能力，树立了工业标准产品发展史上新的里程碑。

对比之下，其他厂商由于首鼠两端或者实力不足都只能推出部分档次的产品，使 HP Integrity 成为当前唯一的基于 IPF 完整的服务器系列。

HP Integrity 系列不仅目前具有全系列产品以及性能和性价比优势，而且还为未来发展有很大的发展空间，并且在扩展过程中保护客户原有的投资。

在处理器方面，HP 的 Integrity 系列基于工业标准的 IPF 系列。Intel 正在投资巨资发展这一系列，从 2001 年的 Itanium 到 2003 年的第二代 Itanium2 性能提高了一倍以上，Intel 将于 2005 年推出双核的 Montecito、2006 年起将推出 4 核到 16 核的 Tukwila 处理器，性能将比 Itanium2 提高 10 倍以上，为 HP 基于 IPF 产品提供了无限的扩展空间(详见[13])。

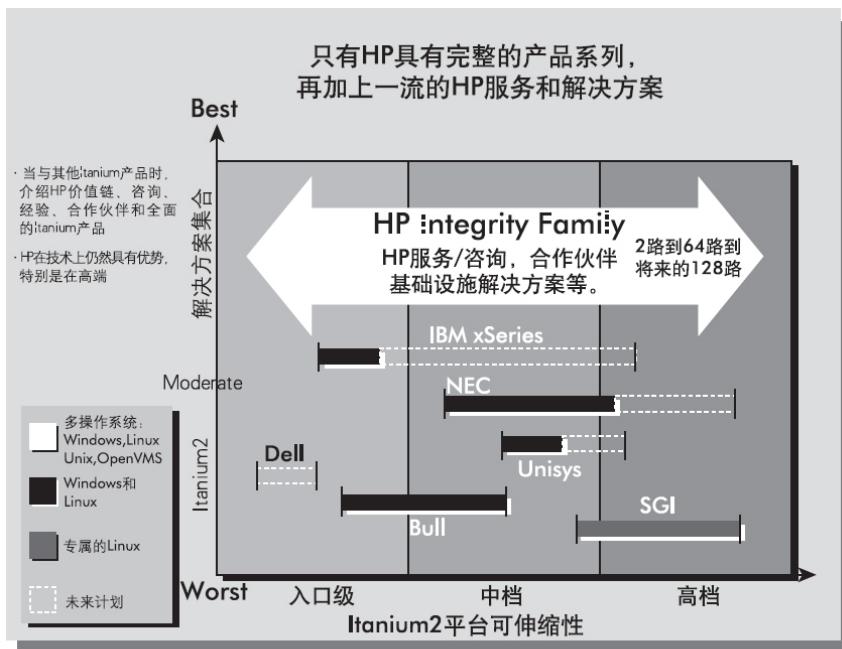


图 1-7 HP Integrity 服务器的覆盖范围与其他 Itanium 厂商相比较

在单机系统中，HP 基于 sx1000 芯片组的 Superdome 目前已经能够支持 64 个第二代 Itanium2 处理器。HP 已经设计一种新型的插件板 mx2，能够插入现有或未来的装备 Itanium2 以后芯片的系统，将处理器的数量增加 1 倍，从而使得 Superdome 能够支持 128 个 Itanium2 处理器，提供更高的性能和性价比。

在集群系统方面，目前已经有不少用户与 HP 合作，利用 HP 现有的产品建立装备上千 Itanium2 处理器的 Linux 集群系统，提供超级计算能力。例如，HP 与美国能源部西北太平洋国立实验室合作建立的超级计算机，将使用 1438 个 Itanium2 处理器、提供超过 10 TFLOPS 的计算能力。该机在 TOP500 中位居第 5，成为世界上最快的超级计算机之一。

由此可见，用户采用 HP 基于 IPF 产品不但能够保护现有硬件投资，而且将使满足未来业务发展需要得到可靠的保障。

表 1-5 用户对服务器的全面应用需求

全面的性能指标	先进的系统特性	厂商能力和水平
<ul style="list-style-type: none"> <li>● 系统基准测试指标</li> <li>● 通用的应用测试指标</li> <li>● 具体的应用测试指标</li> </ul>	<ul style="list-style-type: none"> <li>● 可伸缩性和投资保护</li> <li>● RAS(可靠性、可用性、可维护性)</li> <li>● 可管理性</li> <li>● 总拥有成本</li> </ul>	<ul style="list-style-type: none"> <li>● 全面的解决方案</li> <li>● 丰富的应用软件</li> <li>● 合作伙伴队伍</li> <li>● 技术服务和支持</li> <li>● 财务服务</li> </ul>

表 1-6 zx1 芯片组支持的内存容量和带宽			
指标	直接联接内存	通过内存扩展芯片联接	
		使用 6 个内存扩展芯片	使用 12 个内存扩展芯片
内存容量	16 DIMM	32 DIMM	64 DIMM
内存带宽	8.5 GB/sec	12.8GB/sec	12.8 GB/sec
内存延迟	112 ns	127 ns	127 ns

表 1-7 sx1000 5 种 ASIC 的特性和功能		
ASIC 类型	特性	功能
单元控制器	交叉存取或本地化内存	灵活分配物理地址：跨单元访问各单元上的内存地址
	基于目录的缓存一致性	保证处理器缓存、IO 缓存和主存间的一致性
	ECC	纠正内存和接口上 1 位的内存错误, 探测 2 位错误
	纠正内存和接口上 1 位的内存错误, 探测 2 位错误	纠正由于整个 DRAM 故障而产生的内存错误
内存缓冲区	内置的目录修改电路	提供加速实现基于目录的缓存一致性机制
PCI-X 系统	高速缓存	包含两个 8 KB 缓存, 用以加速 IO 卡 DMA 传输
总线适配器	转换缓存	转换使得 32 位 PCI 卡能够访问所有内存空间 (最大 2 TB)
PCI-X 主机桥	支持热插拔	在线插入或更换 PCI/PCI-X 卡
	错误隔离	包含对一条总线的一个 PCI 故障的隔离，不会影响其他 PCI
交叉交换器	8 路非互锁交叉交换器	在端口之间提供独用的路径, 以大队列 降低信息包的延迟

表 1-8 HP sx1000 芯片组技术		
芯片	晶体管数 (M)	信号针数
单元控制器	41	1466
内存缓冲区	1	248
系统总线适配器	14	612
PCI-X 主机桥	12	170
交叉交换器	21	748

表 1-9 主要厂商基于 Itanium2 的服务器产品系列	
厂商和系列名称	NEC Express/1000 系列 (只有支持 8-32 个处理器的中高档产品)
简介	使用专利的 ccNUMA 芯片组, 交叉交换器架构

	支持多操作系统 (仅在日本支持 HP-UX)		
机型和主要参数	Model 1320 Xc 32 个 Itanium2 1.5 GHz 最大内存 256 GB 8 个硬分区 112 PCI 插槽 Linux , Windows , HP-UX 操作系统	Model 1160 Xc 16 个 Itanium2 1.5 GHz Max 内存 64 GB 4 个硬分区 56 PCI 插槽 Linux, Windows, HP-UX 操作系统	Model 1080 Xc 8 个 Itanium2 1.5 GHz Max 内存 256 GB 2 个硬分区 26 PCI 插槽 Linux, Windows, HP-UX 操作系统
厂商和系列名称	Bull NovaScale 系列 (只有支持 4–8 个处理器的中低档产品)		
简介	使用 Intel E8870 芯片组, 多处理器环境灵活架构 (FAME)		
机型和主要参数	NovaScale 5160 16 个 Itanium2 1.5GHz 最大内存 128 GB 4 个动态硬分区 Linux, Windows, GCOS 7/8 操作系统	NovaScale 5080 8 个 Itanium2 1.5GHz 最大内存 64 GB 2 个动态硬分区 Linux, Windows, GCOS 7/8 操作系统	NovaScale 4040 4 个 Itanium2 1.5GHz 最大内存 32 GB Linux, Windows, GCOS 7/8 操作系统
厂商和系列名称	SGI Altix 3000 系列 (产品基本面向高性能技术计算市场)		
简介	NUMAflex 架构, 主要面向 HPTC 市场		
机型和主要参数	Altix 3700 4–64 Itanium, 1GHz 最大内存 512 GB 11 PCI 插槽 专属的 SGI Linux	Altix 3300 4,8,12 Itanium , 0.9GHz 最大内存 128 GB 11 PCI 插槽 专属的 SGI Linux	
厂商和系列名称	Unisys ES7000 (只有支持 2×16 个处理器的高档产品)		
简介	使用 Intel E8870 芯片组, 两层交叉交换互联拓扑		
机型和主要参数	Orion 430 2×16 Itanium2 1.5GHz 最大内存 64 GB 96 I/O 插槽 Windows , United Linux	Orion 560 2×16 Itanium2 1.5GHz 或 32 个 Xeon 处理器 最大内存 64 GB 96 I/O 插槽 Windows, United Linux	
厂商和系列名称	IBM eServer (只有支持 2–16 个处理器的中低档产品)		
简介	使用 IBM 的 EXA-2 架构和 Summit (X-64) 芯片组		
机型和主要参数	xSeries 382 2 Itanium2 , 1.5 GHz 最大内存 16 GB 3 个 PCI 插槽 Windows 2003 ,	xSeries 445 4 Itanium2 , 1.5 GHz 最大内存 40 GB 18 个 PCI 插槽 Windows 2003, Linux	xSeries 455 16 Itanium2 , 1.5 GHz 最大内存 56 GB 18 个 PCI 插槽

	Linux		Windows 2003 , Linux
厂商和系列名称	Dell PowerEdge (只有支持 2 个处理器的低档产品)		
简介	使用 Intel E8870 芯片组		
机型和主要参数	PowerEdge 5250 2 个 Itanium 2, 1.5 GHz 最大内存 16 GB 3 个 PCI-X 插槽 Windows 2003 , Linux		

表 1-10 工业标准 Integrity 服务器平台上的生态系统建设

层次	产品	说明
业务基础件	SAP R/3 Enterprise.Oracle e-Business Suite, Baan 的 iBaan, PeopleSoft 的 eCenter 等业务基础件	HP 在 Integrity 平台上提供一系列著名的支持企业应用的业务基础件, 为广大 ISV 和企业用户开发满足各行各业专门需要和应用环境的解决方案和应用软件提供了有利条件。
系统基础件	应用服务器	HP 提供 BEA WebLogic Platform、Oracle AS 和 SAP Web 以及 Windows 2003 .NET 应用服务器, 提供全面基于 J2EE 和.NET 两种框架下 Web Services 软件开发功能
	数据库系统	HP 提供 Oracle 9i Release 2.0 以及 SQL Server 2003 数据库和开发工具, 全面的数据库管理功能, 是驱动应用服务器的引擎。开发工具包提供支持在数据库上进行应用开发的工具软件。
传统开发工具	编译语言和开发工具	HP 与领先厂商合作提供齐全的编译程序和开发工具, 包括支持并行计算的 MPI 接口和专门优化的 MLIB 子程序库
操作系统	HP-UX 11i, OpenVMS, Linux 和 Windows 2003 操作系统	HP 为用户提供最优秀的操作系统平台和最广泛的选择余地
硬件平台	基于 IPF 的 Integrity 完整的服务器产品系列	HP 已经制订了明确的战略把所有 64 位服务器产品过渡到基于和 IA-64 工业标准架构下的 64 位 Integrity 系列

### 1.4.3 HP 在 IPF 平台上的多操作系统战略

在今天的 IT 环境中, 服务器一般都是为支持一种操作系统设计的。Linux 普及后, 许多服务器开始能够支持两个操作系统。既然如此, 为什么不能支持 3 个甚至 4 个操作系统? 为

什么不能把多个操作系统(如 UNIX, Linux, Windows Server 2003 和 OpenVMS)组合到一个服务器系统上? 事实上, 可靠和高效地支持多种操作系统已经成为服务器技术未来的发展趋势。HP 在 IPF 平台上执行多操作系统战略。这一战略的核心要素是在基于 IPF 的公共平台上、提供可供客户灵活选择的多种操作系统 HPUX 11i, OpenVMS, Linux, 和 Windows 2003(将来还包括 NonStop Kernel)以及跨越所有这些操作系统的公共软件基础设施, 满足企业用户在复杂的异构环境中的需求, 支持 IPF 发展成为高端的主流平台。IT 机构购买能够支持多个操作系统的服务器如 HP Integrity 服务器有很多好处(详见[15]), 包括:

- 增加整合的机会;
- 优化资源利用;
- 便于操作;
- 增加选择应用软件的灵活性;
- 避免风险(保证不会由于改变 OS 而受到损失);
- 便于把硬件移作它用;
- 得到更多的 ISV 支持(提供多种开发/测试软件的环境);

#### 1.4.4 HP 为 Integrity 系列建立最完整的企业应用环境

目前, 高可用性的概念正在被连续可用性(或业务连续性)所代替。企业真正关心的不仅是系统正常运行, 而是保持系统所支持的业务正常运行、及时提供高质量的结果、使客户在任何时间得到所需的信息和服务。因此, 业务连续性比传统的高可用性前进了一步。连续可用性有 5 个组成部分: 高可靠性、高可用性、灾难恢复功能、服务水平管理和消除手工管理错误。

基于 Itanium 的产品作为一种新产品是否具有高可用性、以至连续可用性, 关系到它们是否能够应用于支持关键任务应用, 也关系到 Itanium 的前途。HP 基于 Itanium 产品全面提供实现连续可用性的 5 种特性, 支持完整的业务连续性解决方案, 使得 Itanium 和基于它的产品能够顺利进入关键任务系统行列, 满足企业日益增长的需求。

传统的 Linux 系统的可用性、安全保密和管理功能低于 UNIX 系统, 使企业用户往往不敢把它应用于支持关键任务, 成为 Linux 在企业中推广应用主要障碍。HP 通过为工业标准平台上的 Linux 提供高可用性和容灾能力、自动和智能的管理和严格的安全保密功能, 建立适合于运行企业关键任务应用的 Linux 环境, 有力地推动 Linux 在企业中的应用, 特别是在金融服务、银行和电信等对可用性、安全性和可管理性要求最高的行业中的应用。

在提供高可用性和容灾能力方面, HP-UX 上的 MC/ServiceGuard 集群软件是最著名和成熟的高可用性集群软件, 已经售出了 7 万个许可证。HP 把它在高可用性应用领域丰富的经验应用到 Linux 环境, 把 MC/ServiceGuard 移植到 Linux 上, 并增加了支持 8 个节点的自动故障切换, 提供基于 Linux 高可用性产品解决方案。该方案还把 Linux 操作系统的 MC/ServiceGuard 软件与使用 HP StorageWorks XP 磁盘阵列的 StorageWorks Cluster Extension 集成在一起, 保护地理上分开的数据中心(最大距离可达 100 公里), 防止由于系统和应用故障、操作员误操作或者自然灾害引起服务中断。应用于 HP StorageWorks NAS8000 的嵌入高可用性解决方案把 MC/ServiceGuard for Linux Clustering 与 NAS8000 管理功能组合在一起、提供高可用的 NAS 解决方案。此外, HP 还提供 SteelEye 公司的 Linux 集群软件 LifeKeeper、支持建立基于 HP ProLiant 服务器和 StorageWorks 的高可用性数据集群解决方案。

在提供高可管理性方面, HP 把领先的基础设施管理平台 OpenView 移植到 Linux 上提供智能和自动的管理解决方案。OpenView 通过合作伙伴程序支持管理大量的应用软件, 如

通过 Smart Plug-Ins 支持在 Linux 上运行的 Oracle 和 SAP。HP OpenView 是一个全面的管理解决方案，能够监控网络、系统、存储、应用软件、数据库和服务。HP OpenView Operations Application 具有能够监控 Linux 系统运行状态和性能的代理软件。HP OpenView Network Node Manager 能够发现 Linux 设备；HP OpenView Internet Services 监控基于 Linux 的服务。HP OpenView Omniback II 为 Linux 系统提供数据备份和恢复保护功能。通过把 OpenView 移植到 Linux，HP 在基于 Linux 的系统上建立了领先的企业管理功能；通过移植 ServiceControl Manager，HP 为 Linux 集群系统提供单点管理功能。这一可管理性工具提供多系统管理功能，如成组操作、基于分工管理等，在执行任何管理任务之前都要进行用户确认，保证整 IT 环境帐户管理和收费的安全可靠；HP 通过移植 Process Resource Manager 为 Linux 系统提供 CPU 资源管理功能，允许系统管理员监测、控制和优化系统资源。

2003 年 11 月 13 日 HP 推出跨 Windows, Linux 和 Unix 服务器产品线系统管理功能的新软件。这个跨平台产品名为 Systems Insight Manager (内部代号 Nimbus)，把 HP 以前分别销售的 ServiceControl Manager、Insight Manager 7 以及 Tooptools 等三个工具优点和功能组合在一起，统一管理公司的 ProLiant, Integrity 和 HP 9000 服务器。这一工具也处理管理 PC、存储设备、打印机和电源以及某些第三方软件的插入工具。此外，Systems Insight Manager 可以与 HP 的旗舰企业管理工具 OpenView Operations 以及 OpenView Network Node Manager 集成在一起，提供更强管理功能。

在安全保密方面，HP 提供的 HP Secure Linux 通过支持防止入侵、实时防攻击保护和 damage containment 等功能，帮助企业保护它们的 Linux 环境。HP 是在市场上销售 Linux 关键任务安全解决方案的唯一系统厂商。新产品包括与 Sendmail, ftp, DNS, LDAP, Apache, Tomcat, Struts, NFS, SNMP, and Samba 等关键子系统集成的版本；

#### 1.4.5 HP 在 Integrity 系列平台上建立强大的生态系统

在生态系统建设方面，传统的基于“操作系统 + 计算机语言和开发工具”或者基于“操作系统 + 中间件”的生态系统架构已经逐步为基于“操作系统 + 系统基础件 + 业务基础件”架构所取代。在这一架构中：

- 操作系统也可以称为操作系统平台是最底层和基础性的软件，提供设备管理、作业调度和人机对话等功能，使上层的软件不必考虑硬件系统的具体细节；
- 系统基础件的底层是数据库软件，核心是应用服务器，外围是门户、集成、系统管理、安全和开发等相关的部件。系统基础件提供一个通用的平台，使上层的软件开发不必考虑操作系统的许多细节，大大提高了上层软件通用性、可移植性和开发效率；
- 业务基础件是新出现和最有潜力的一层。它构筑于基础件之上、为开发各种应用解决方案提供基础架构。它提供管理应用软件的开发和运行、与操作系统、基础件之间的交互等功能。同时也屏蔽了下层软件的技术细节，使开发人员能够全力解决软件研发中的具体业务和管理问题，摆脱技术细节的困扰、从而大大提高软件的开发效率；

HP 与 Microsoft, Oracle, BEA 等领先厂商合作在工业标准的 Integrity 服务器平台上所有操作系统下，提供先进和功能强大的编译语言、开发工具、系统基础件和业务基础件，建立了完整的生态系统，大大提供了广大 ISV 和用户在 Integrity 所有操作环境中应用开发的效率、提高了 IPF 的通用性，为基于 IPF 产品在企业中推广应用创造了有利的条件、促进了 IPF 成为支持高端应用的主流平台。

#### 1.4.6 HP 为 Integrity 提供最丰富的解决方案和应用软件

64 位 RISC 在 10 余年中积累了大量的解决方案和应用软件，确保了基于 64 位 RISC 处理器的服务器在几乎一切领域中的应用。基于 IPF 服务器问世之初，解决方案和应用软件太少，成为阻碍其推广应用的主要障碍之一。HP 为 Integrity 系列提供最丰富的应用解决方案，推动 Integrity 服务器广泛应用，成为 HP Integrity 系列的最重要优势之一。HP 在 Integrity 平台上提供包括 CRM、BI、ERP、SCM 和 HPTC 等跨行业的解决方案以及面向金融、电信、制造等许多具体行业的解决方案，应用软件数量也已经接近 1000 个，满足当前最紧迫的应用需求(见[14])，到 2004 年中应用软件数量将超过 2000 个。

#### 1.4.7 HP 为 Integrity 系列提供最全面的服务和支持

随着企业 IT 系统规模的扩大、结构的复杂化以及可用性要求的提高，技术服务的重要性也与日俱增。目前，人们已经把服务和支持与硬件和软件并列为企业 IT 基础设施的组成部分之一。HP 正以其在实力和优势，在 IPF 平台上提供周全的技术服务和支持。

HP 全球服务部门是世界上最大 IT 服务部门之一，在 160 多个国家中拥有 65000 多名服务专业人员，包括 28000 多名经微软培训的专家、5000 多名通过微软认证的工程师、18000 多名 UNIX 专家、3000 多名 Linux 专家。HP 实力雄厚的服务部门在 IPF 平台上提供跨越 Windows、UNIX 和 Linux 等三个操作系统的全面技术服务和支持，包括：

- 战略咨询服务
- 移植和迁移服务
- 实施和集成服务
- 支持服务，包括硬件、软件和关键任务服务以及远程支持技术服务；
- 培训服务
- 金融服务

此外，HP 投入大量资金在亚太地区的北京、东京、新加坡、汉城、Bangalor 和悉尼等地建立了合作伙伴技术支持中心(PTAC)，并在台湾和香港建立子中心。PTAC 网络为 Windows、Linux 和 Unix 的开发人员提供技术帮助，使 32 位系统应用或基于其他 64 位平台的应用向 IA-64 实现平滑过渡。PTAC 提供的服务项目包括：

- **IPF 评估服务：**PTAC 的咨询顾问将使用优化的过渡策略，帮助开发人员最充分地利用其有限的技术资源，使所开发的应用程序迅速在 IPF 系统上能够运行；
- **过渡服务：**PTAC 与 ISV 开发人员合作，使用跨平台编译工具或 IA-64 的开发工具把 HP-UX、Windows 和 Linux 应用程序过渡到 HP 的 IPF 平台系统上。今后，HP 还为 Tru64 UNIX 和 OpenVMS 应用程序提供过渡服务；
- **认证服务：**PTAC 提供全面的基础结构，以帮助开发人员在 IPF 架构上运行的 HP-UX 应用程序通过认证；

这些服务和支持大大丰富 IPF 平台上应用软件的阵容，并且已经取得了很好的效果。目前全球已有近 1000 个基于 IPF 平台的应用软件可以投入正式运行。

## 二、HP Integrity 系列入口级服务器

入口级服务器的一般特征是：装备 1-4 个 CPU、价格不超过 10 万美元、在网络边缘或工作组层次应用。HP 基于 Itanium2 的 Integrity 系列目前包括 rx2600、rx4640 和 rx5670 等三款入口级服务器。HP 通过全面贯彻其在工业标准部件基础上进行增值的战略，使 HP 基

于 Itanium2 的入口级服务器具有最高的性能、性价比、兼容性和可靠性、最低的入口价位和总拥有成本最低、最小的体积、最广泛的操作系统选择余地和最全面的解决方案，提供领先于其他厂商基于 Itanium2、Opteron、IA-32 和 RISC 处理器入口级服务器的竞争优势。本章介绍 HP Integrity 系列入口级服务器的产品概貌、架构、特性和竞争优势。

## 2.1 HP Integrity 系列入口级服务器概述

HP Integrity 系列提供 rx2600、rx4640 和 rx5760 等型号的入口级服务器，其硬件的基本参数如表 2-1 所示。

## 2.2 HP Integrity 系列入口级服务器架构

HP Integrity 系列入口级服务器采用对称多处理器(SMP)架构、交叉交换互联拓扑。这一架构是基于 HP 专门设计的 zx1 芯片组(见 1.3 节)实施的，奠定了全面领先优势的基础。

### rx2600 服务器架构

HP Integrity rx2600 服务器支持 1 或 2 个 Intel Itanium 2 处理器，通过一条 200 MHz 双转储 128-bit 前端系统总线联接到 HP zx1 芯片组内存和 I/O 控制器。系统总线上的总带宽为 6.4 GB/s。

内存 DIMMs 直接联接到 2 条 266 MHz, 4.3 GB/s 内存总线。两条总线的组合内存带宽为 8.5 GB/s。每条总线联接到最多 6 个双倍速率(DDR) SynchDRAM 内存 DIMMs。通过 12 个 2 GB DIMM 提供 24 GB 内存总容量。

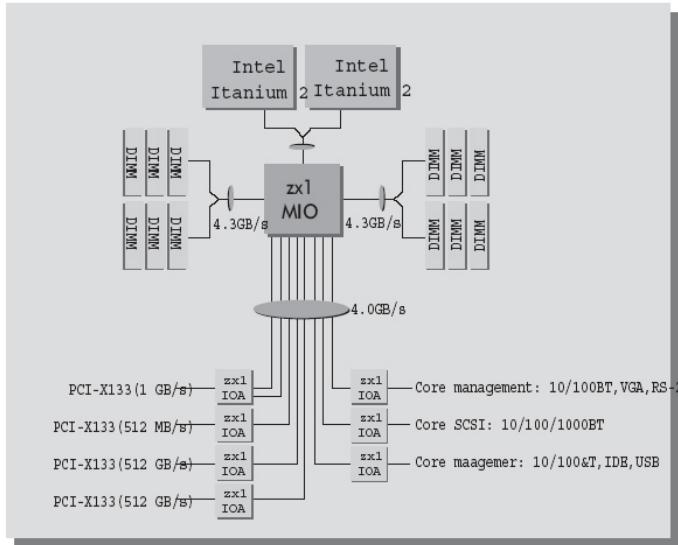


图 2-1 Hp Integrity rx2600 架构

I/O 架构由 8 个 0.5 GB/s 通道组成，分配到 7 个 zx1 I/O 适配器。每个适配器提供一条与系统中 I/O 插槽或核心 I/O 设备相联接的 PCI-X 或 PCI 总线。前 2 个 I/O 通道联接 1 个 133 MHz PCI-X I/O 插槽，提供 1 GB/s 持续吞吐能力。这一插槽特别适合于高带宽 I/O 适配器，如高性能集群互联。其后的 3 个 I/O 通道联接到 3 个独立的 133 MHz PCI-X I/O 插

槽，每个提供 0.5 GB/s 持续吞吐能力。其余 3 个 I/O 联接到 3 条 PCI 总线，它们又建立到核心 LAN、SCSI、IDE、USB 接口和管理处理器的联接。

## rx4640 服务器架构

HP Integrity rx4640 服务器支持 1、2、3 或 4 个 Intel Itanium 2 处理器，通过一条 200 MHz 双转储 128-bit 前端系统总线联接到 HP zx1 芯片组内存和 I/O 控制器。系统总线上的总带宽为 6.4 GB/s。

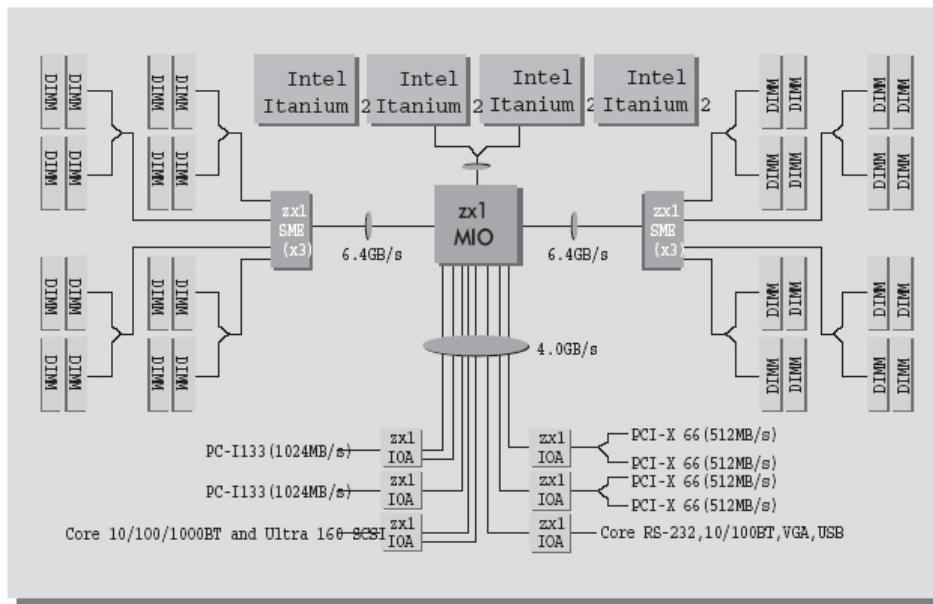


图 2-2 Hp Integrity rx4640 服务器架构

zx1 内存控制器联接到两个独立的 200 MHz, 6.4 GB/s 内存总线。每条总线联接 6 个 zx1 可伸缩内存扩展器，把带宽分配给双倍速率(DDR)SynchDRAM 内存 DIMM。DIMM 总数为 16 或 32 个单元，分布在 1 块 内存板上，提供最大 64GB 的内存容量。

I/O 架构由 8 个 0.5 GB/s 通道组成，分别与 6 个 zx1 I/O 适配器相联接。每个适配器提供一条与系统中 I/O 插槽或核心 I/O 设备相联接的 PCI-X 或 PCI 总线。前 2 个 I/O 通道联接 1 个独立的 133 MHz PCI-X I/O 插槽，提供 1.0 GB/s 的持续吞吐能力。其后的 2 个 I/O 通道联接到一个相同的 133 MHz PCI-X I/O 插槽；后 2 个 I/O 通道联接到一对 zx1 I/O 适配器，每一个联接到一对 66 MHz PCI-X I/O 插槽。每对插槽共享 0.5 GB/s 带宽。最后两个 I/O 通道联接到核心 I/O。一个通道提供 0.5 GB/s 带宽联接到核心 10/100/1000 BT LAN 和双通道 Ultra160 SCSI 控制器；另一个通道提供 0.5 GB/s 带宽联接到核心管理 LAN、RS-232 串行端口、USB 端口和 VGA。

## 5670 服务器架构

HP Integrity rx5670 服务器支持 1、2、3 或 4 个 Intel Itanium 2 处理器，通过一条 200 MHz 双转储 128-bit 前端系统总线联接到 HP zx1 芯片组内存和 I/O 控制器。系统总线上

的总带宽为 6.4 GB/s。

zx1 内存控制器联接到两个独立的 200 MHz, 6.4 GB/s 内存总线。每条总线联接 6 个 zx1 可伸缩内存扩展器，把带宽分配给双倍速率(DDR)SynchDRAM 内存 DIMM。总的 DIMM 容量是 48 个单元，分布在 2 块 24 个 DIMM 内存承载板上，每条内存总线 1 块板。但是，服务器只能支持 24 个 DIMM 单元的内存容量。因此，如果把 24 个 DIMM 全部放置在一块内存板上，将只能利用一半内存带宽。为了最大限度提高性能，rx5670 应当配置成把所有 DIMM 单元均分在 2 块内存板上，从而完全使用 rx5670 的 12.8 GB/s 内存带宽。

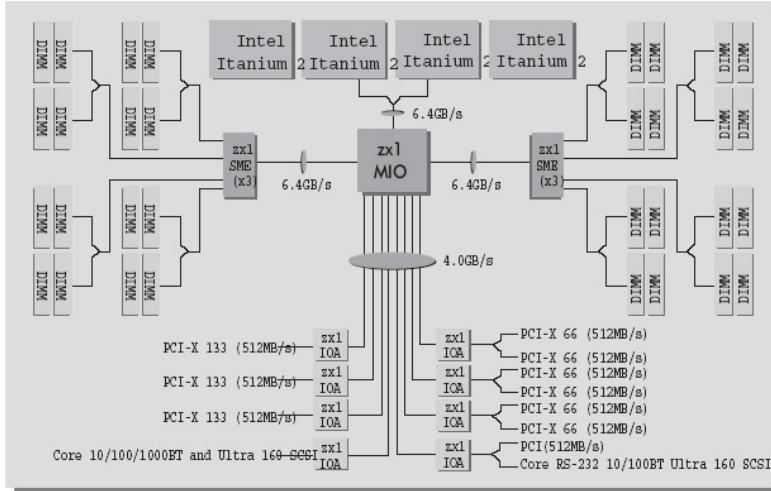


图 2-3 Integrity rx5670 架构

I/O 架构由 8 个 0.5 GB/s 通道组成，分别与 8 个 zx1 I/O 适配器相联接。每个适配器提供一条与系统中 I/O 插槽或核心 I/O 设备相联接的 PCI-X 或 PCI 总线。前 3 个 I/O 联接 3 个独立的 133 MHz PCI-X I/O 插槽，每个有 0.5 GB/s 持续吞吐能力。其后的 3 个 I/O 通道联接到 3 个 zx1 I/O 适配器，每一个联接到一对 66 MHz PCI-X I/O 插槽。每对插槽共享 0.5 GB/s 带宽。

最后 2 个 I/O 通道联接到核心 I/O。一个通道提供与核心 10/100/1000BT LAN 以及一个双通道 Ultra160 SCSI 控制器相联接的 0.5 GB/s 带宽。另一个通道提供与核心管理 LAN, RS-232 串行端口、另一个双通道 Ultra160 SCSI 控制器以及一个 33 MHz PCI 插槽相联接的 0.5 GB/s 带宽。33 MHz 插槽特别适合于可选的 VGA/USB 适配器。在工厂预装 Microsoft Windows 操作系统的系统上总是把 VGA/USB 卡放置在这一插槽上。

## 2.3 HP Integrity 系列入口级服务器特性

HP Integrity 系列入口级服务器具有超过一般 32 位 PC 服务器的 RAS 和可管理特性，满足企业应用的需求。

### 2.3.1 RAS(可靠性、高可用性和可维护性)特性

HP Integrity 入口级服务器是按照支持关键任务环境的需要来设计的，能够提供 99.95% 到 99.999% 的可用性(取决于解决方案的具体配置)。这要求服务器硬件本身具有很高的单系统高可用性(SSHA)。HP Integrity 入口级服务器的设计从机架基础设施、I/O 子系统到内存和处理器子系统都充分考虑了高可用性的需求，使整个系统本质上具有内置冗余性和自愈功能。HP 的故障事件监控服务又进一步增强了它们的 SSHA。为了实现最高的可用性，可以利用集群软件(如 HP 的 Serviceguard)把它们联接成集群系统。

## **高可用性机架基础设施 (电源和冷却设备)**

rx5670, rx4640 和 rx2600 上的风扇从设备前面抽风、通过内部的系统部件把热量散发到服务器后面，提供非常好的致冷功能。所有风扇都很容易装修，提供 N+1 冗余度。它们的电源子系统也提供 N+1 冗余电源选购件。rx5670 的标准配置包含两个热交换的电源，提供可选购第三个电源、可以实现 2+1 冗余。rx4640 和 rx2600 的标准配置包含 1 个热交换的电源，提供可选购第二个电源、可以实现 1+1 冗余。为了进一步增强可用性，每个电源都有自己专用的电源输入线，可以插入不同的电网中、提供最高水平的电源保护。

## **磁盘驱动器热插拔和磁盘镜像**

rx5670, rx4640 和 rx2600 所有磁盘都可以从系统前面装修，并在服务器运行过程中插拔。在 rx5670 中，使用两个双通道 SCSI 控制器管理 4 个热插拔磁盘。为了增加可用性，磁盘对可以装在不同的通道和不同的 SCSI 控制器上。这意味着利用磁盘镜像功能，当一个 SCSI 控制器、一个 SCSI 通道或根磁盘出故障时，系统仍然能够正常运行。在 rx4640 中，使用一个双通道 SCSI 控制器管理磁盘对。两个磁盘可以配置在一个 SCSI 通道上、或者每个通道上配置一个磁盘、建立磁盘镜像，提高可用性。在 rx2600 中，使用一个双通道 SCSI 控制器管理 3 个磁盘。一个通道联接 2 个外部磁盘，另一个通道联接第三个外部磁盘。这允许跨不同的 SCSI 通道建立磁盘镜像、进一步增强可用性。

## **多 I/O 通道**

rx5670, rx4640 和 rx2600 中的多 zx1 I/O 通道提供故障恢复、负载均衡和故障隔离功能。在这些服务器中，一个通道上的故障不会干扰其他通道的操作。此外，服务器还提供完全独立的 PCI-X 总线来隔离 I/O 适配器上的信息传输。如果一个适配器上产生故障，它将不会干扰另一个总线上的信息传输。

## **ECC 和芯片备份内存**

HP Integrity rx5670, rx4640 和 rx2600 服务器的内存子系统利用错误校正代码纠正单位错误、利用 HP 的芯片备份技术防止发生多位错误。芯片备份技术使用 DIMM 上的整个 SDRAM 芯片在发生多位错误时被旁通。为了使用芯片备份功能，必须以四元方式装入相同容量的 DIMM。系统也在不同的四元组中支持不同容量的。rx5670, rx4640 和 rx2600 服务器利用芯片备份特性实现在发生 SDRAM 故障时完全自愈，而不管故障有多少位出故障。这将最终避免内存成为系统的故障来源。

## **CPU 错误校正和动态处理器自愈**

在 HP Integrity rx2600, rx4640 和 rx5670 服务器中，L1 和 L2 缓存都有完全检测和校正 1 位错误以及探测 2 位错误的功能。此外，所有指令和数据传输路径都具有检测和校正 1 位错误的功能。系统处理器总线还具有探测奇偶错误功能，数据路径具有错误校正功能。这些服务器也都能够实现动态处理器自愈(DPR)。利用 DPR，任何以不可接受速率产生可探测缓存错误的 CPU 将被从系统去分配(操作系统将不再把新的进程调度到其上)。这一特性有助于防止 CPU 可靠性下降到可能成为系统崩溃的来源。

## 全面的错误记录

所有系统事件被存储在永久存储的系统事件记录 (SEL) 中。此外，系统半固件在永久存储中创立一个活动和前推记录。除了在非常极端的情况下，这些信息将足以把系统故障诊断到单个可置换的部件。SEL 和 FPL 可供管理处理器(因此可以远程调用)和系统级工具使用，实现快速和精确诊断。

## 整个生命期的故障管理

故障管理是 HP 的全面战略和实施计划、提供完全的价值链探测、通知和维修系统问题。故障管理从设计阶段就开始实施，此时硬件和 OS 设计人员加入了探测和隔离系统异常的功能和指令点。也设计了监控软件获取系统的健康信息或异步响应应到预先设计在系统中的指令点、报告问题或者故障。故障管理也涉及建立几种维护历史事件信息的方法、允许保留供分析和发现系统变化趋势用的信息。产生错误和警告的故障自动记录到 syslog，同时把通知和审计信息复写到事件记录中，也可以选择地保留历史信息。故障管理在探测到问题后(即使是潜在的问题)提供立即的警告，使用户能够采取正确定行动。在某些情况下，故障监控软件甚至能够维修系统、防止故障重新发生。

## 故障监控软件的功能

障管理与监控功能相结合能够建立系统部件运行状况表、产生接近实时的事件探测功能。这些事件能够触发系统采取正确定行动、保持连续的正常运行，也能够发警报给系统管理人员，使他们能够在问题变严重前加以处理。故障监控软件能够：

- 获取系统的运行状态信息；
- 处理已经设计在硬软件中的异步事件；
- 执行可能的校正行动；
- 去分配正在出问题的内存以免产生故障(动态内存自愈)；
- 去分配正在出问题的处理器以免产生故障(动态处理器自愈)；
- 在下次自举前，把出故障的处理器清除出现有的配置；
- 当电源故障造成转而使用 UPS 时关闭系统；
- 管理事件、使系统性能不会由于错误而下降；

- 提供有关问题原因和应采取的行动的信息;

## 通知和一体化企业管理

故障管理当前使用 HP EMS (事件监控服务)基础设施来完成通知任务。EMS 使用广泛类型的通知方法，包括页面、e-mail、SNMP 服务、系统控制台、字符记录文件 TCP/UDP 和 OpenView 操作中心(OPC) 消息。故障管理事件可以在服务器上直接观察、或者通过 HP Insight Manager 观察，也可以是数据中心中多个系统的合并信息。用户也可以选择把故障管理事件与 HP (OpenView)或 BMC, Tivoli, Computer Associates, 或 MicroMuse 公司的企业管理软件集成在一起。

### 2.3.2 高可管理性

HP Integrity 系列入口级服务器通过硬软件结合的方法、提供最高的可管理性。

#### 可延伸的半固件接口

可延伸的半固件接口(EFI) 是 HP-UX, Linux 和 Windows 操作系统与基于 Itanium 2 平台半固件之间的接口。可延伸的半固件接口支持的文件系统是基于文件分配表 (FAT) 的文件系统。EFI 允许使用 FAT-32 作系统分区。系统分区是一组连续的磁盘分区，是基于 Itanium2 平台的自举盘上必须具备的。

基板管理控制软件提供使系统便于管理的功能。基板管理控制软件支持工业标准的智能平台管理接口(IPMI) 规范。这一规范描述已经置入系统中的管理特性，包括诊断、配置管理、硬件管理和故障检测。基板管理控制软件与管理处理器联接提供最高水平的系统可管理性和高可用性监控功能。

基板管理控制器提供如下的功能：

- 电源和复位管理；
- 系统运行状况管理：风扇、电源、温度和电压；
- 事件记录和报告：系统事件记录、前推记录、状态板上的诊断 LED；
- 设备联接配置；
- 硬件和数据保护：在发生关键事件时自动实现安全的 OS 关闭、保护存储的系统配置参数和系统启动 ROM；
- 通过 IPMB 联接到专门的外置管理处理器(MP)、实现通过 MP LAN 或 MP 串行端口作远程管理；
- 与智能平台管理接口 1.0 兼容；

#### HP 管理处理器

管理处理器是 rx2600, rx4640 和 rx5670 服务器的标准组成部件。该处理器提供到基板管理控制软件的远程接口、管理系统资源、诊断系统的运行状况和进行系统维修。管理人员

可以通过专门的带外(即独立于系统主数据路径的)通信联线与管理处理器对话。

管理处理器能够最大限度减少以至消除系统管理员亲身在系统上执行诊断、系统管理和硬件复位等操作的需要。rx2600, rx4640 和 rx5670 管理处理器的主要特性是：

- 在 Internet or intranet 网上进行系统管理；
- 系统控制台重新导向；
- 控制器镜像；
- 支持自动重新启动的系统配置；
- 观察系统事件的历史记录和控制台活动的历史记录；
- 设置 MP 不活动超时阈值；
- 远程系统控制和远程电源状态管理；
- 观察系统状态；
- 事件通知系统控制台、e-mail、pager 和/或 HP Response Centers, e-mail 和页面显示器通知与 HP Event Monitoring Service (EMS)共同完成；
- 关键环境问题的自动硬件保护；
- 在 WAN 故障时访问管理接口和控制台 (需要 modem)；
- 自动系统重新启动；
- 管理前推指示器(通过虚拟前面板)；
- 带外可管理性和系统半固件升级；
- 可管理性和控制台安全性配置；
- 提供 Web 控制台访问的安全插座层(SSL)密码；

## 2.4 HP Integrity 系列入口级服务器竞争优势

Itanium2 领先的性能和 HP 先进的系统技术奠定了 HP Integrity 系列入口级服务器竞争优势的坚实基础。本节从分析入口级服务器市场竞争势态出发，说明 HP Integrity 入口级服务器不仅具有领先的基准测试指标(SPEC 和 SPEC Rate)、在许多重要和代表性浮点计算(如 Linpack)和企业应用(如 SPECwebSSL, SPECjbb, TPC-C, SAP 和 Oracle 应用服务器等)的性能指标都领先于竞争对手，而且具有领先性价比和最低的总拥有成本，成为开展计算密集的 HPTC 应用、类型广泛的企业管理应用的最佳选择。

### 2.4.1 市场竞争势态

入口级服务器市场竞争势态十分复杂，其焦点是如何在资源和价格的限制条件下，使产品具有最高的性能、性价比、兼容性和可靠性、最低的入口价位和最低的总拥有成本、最小的体积，从而夺取最大的工作组层次应用的市场份额。过去 Intel IA-32 体系结构处理器在服务器市场中份额很小。Xeon 的成功使得 Intel 在服务器市场中低端开始占有越来越重要的地位。特别是在低端夺取了 RISC/UNIX 服务器相当大的市场份额。因此，HP Integrity 系列入口级服务器在市场上不仅需要面对其他厂商基于 Itanium2 和 RISC 处理器的入口级服务器的竞争，而且需要面对基于 Wintel 工业标准的 32 位 Xeon/Xeon MP 服务器的竞争。此外，2003 年 4 月 AMD 公司推出了 Opteron 处理器，基于 Opteron 的服务器也成为入口级竞争场的一个新手。

表 2-4 描述基于 Itanium2 的 Integrity 入口级服务器与基于其他体系结构处理器相应档次服务器的总体竞争势态。

### 2.4.2 工业标准平台上的技术优势

HP 在工业标准技术基础上进行创新和增值的产品总体战略，奠定了 HP Integrity 系列的技术优势，确保该系列的服务器产品不仅提供全面领先的性能、最佳的价格/性能，而且具有一系列有利于支持企业应用的关键特性。一方面，HP 是工业标准处理器平台最积极和有力的支持者，制订了明确的战略和发展蓝图、把高端服务器产品全面转向 Intel Itanium 体系结构的 IPF 处理器系列。HP 使用工业标准 Itanium2 处理器、模块化设计和标准化部件为服务器产品提供领先性能和性价比奠定了坚实的物质基础；另一方面，HP 对基于工业标准的服务器产品进行了一系列创新和增值，包括采用适合系统资源规模和应用的体系结构、互联拓扑，基于 HP 独特的芯片组最能够发挥硬件潜力的软件进行系统实施，为用户提供最广泛的操作系统选择空间，开发丰富的系统和应用解决方案等，从而确立了 HP 在工业标准平台上的技术优势，奠定了 HP Integrity 系列各个档次服务器在性能、性价比和企业应用特性等方面的竞争优势和市场领先地位。

### 性能领先的 Itanium 2 服务器

HP Integrity 系列服务器所基于的 1.5/1.4 GHz Intel Itanium2 处理器是当前速度最快的 64 位处理器，提供 64 位寻址能力以及最高的整数和浮点基准测试指标，为 Integrity 系列入口级服务器提供领先于基于 RISC 和 IA-32 处理器的性能奠定了坚实的基础。从发展趋势看，Intel IPF 处理器系列性能提高的速度也将高于其他处理器系列，从而确保 HP 基于 IPF 的服务器产品的长期领先地位(详见[13])。

### HP 独特的 zx1 芯片组

当前许多公司如 HP、Intel、日立、IBM 和 NEC 等都提供支持安腾 2 的芯片组。其中，HP 的 zx1 芯片组是专门为支持 1-4 处理器的服务器(或 1-2 处理器的工作站)设计的，不仅价格便宜，而且在内存带宽和延迟等方面有明显的优势，奠定了 HP 基于 Itanium2 服务器提供领先于其他厂商基于 Itanium2 服务器的性能和性价比的基础。表 2-5 列出 HP 的 zx1 与 IBM 的 Summit 芯片组和被众多 OEM 厂商采用的 Intel E8870 芯片组主要参数的比较结果。

HP zx1 芯片组提供远比 Intel E8870 和 IBM Summit 芯片组低内存延迟和传输速度，并且避免了支持 NUMA 体系结构的不必要开支，使得 HP 围绕 zx1 芯片组建立的入口级服务器提供比 IBM 和 Dell 基于同样处理器的服务器高的性能和性价比。

#### 2.4.3 提供领先的性能

HP Integrity 入口级服务器提供全面领先的基准测试指标和应用性能，满足入口级服务器所针对的市场需求。

#### 全面领先的性能

HP Itanium 技术的优势确保其基于 Itanium2 入口级服务器在主要的企业应用领域提供全面领先于竞争对手同档次的系统，有力地促进了 Itanium 向广泛的市场领域迅速扩展。表

2-7 说明基于 1.5 GHz Itanium2 的 rx2600, rx5670 服务器已经成为多种操作系统下性能最高的 2, 4 处理器服务器, 具有全面的领先地位。

表 2-1 HP Integrity 系列入口级服务器基本参数			
	rx2600	rx4640	rx5670
处理器 个数类型	1 或 2 个主频为 1.3GHz 或 1.5GHz Itanium2 处理器	1,2,3 或 4 个主频为 1.3GHz 或 1.5GHz Itanium2 处理器	1,2,3 或 4 个主频为 1.3GHz 或 1.5GHz Itanium2 处理器
内存容量	1-24 GB	1-64 GB	1-96 GB
内存带宽	8.5 GB /s	12.8 GB/s	12.8 GB/s
芯片组	HP zx1 芯片组		
扩展插槽	1 个 PCI-X 插槽, 1 GB/s 持续带宽, 64- 位 133MHz 3 个 PCI-X 插槽, 0.5GB/s 持续带宽, 64-位 133MHz 每个插槽 都是全长度的, 有独 立的总线	2 个 PCI-X 插槽, 在独 立总线上, 1 GB/s 持 续带宽 64-位 133MHz 4 个 PCI-X 插槽, 在 2 条共享总线上, 0.5GB/s 持续带宽, 64-位 66MHz	3 个 PCI-X 插槽, 在独立 总线上, 64-位 133MHz, 1 GB/s 持 续带宽 6 个 PCI-X 插槽, 在 3 条共享总线上, 64- 位 66MHz, 0.5GB/s 持 续带宽, 1 个 PCI 插槽 用于图形/USB, 64-位 66 MHz, 0.5GB/s 持 续带宽
热插拔磁盘 驱动器	3 个港湾, 3.5 英寸 磁盘, 438 GB 最大内 部存储, 集成的双通 道 Ultra320 SCSI 控 制器	2 个港湾, 3.5 英寸磁 盘, 292 GB 最大内部存 储, 集成的双通道 Ultra160 SCSI 控制器	4 个港湾, 3.5 英寸磁 盘, 584 GB 最大内部存 储, 集成的双通道 Ultra160 SCSI 控制器
可移动介质	1 个滑轨介质港湾, 可选择 16X-ROM 或 16X/10X/40X CD-RW	可选择 DVD-ROM 或 DVD (与 CD 写兼容)	1 个供可选 SCSI 设备使 用的港湾, 可以选择 DVD-ROM 或 DDS-3
核心互联端口	Gigabit-TX LAN , 10/100 BT LAN , Ultra320 SCSI, 2 个 通用的 RS-232 串行端 口, VGA, 4 个 USB 端口	Gigabit-TX LAN , 10/100 BT LAN , Ultra160 SCSI, VGA, 2 个 USB 端口	Gigabit-TX LAN , 10/100 BT LAN , Ultra160 SCSI, VGA, 2 个 USB 端口
管理处理器互联	10/100 BT 管理 LAN (Web 控制台访问) RS-232 本地控制台, RS-232 远程/Modem 控制台, RS-232 通用		
高度	2 U	4 U	7 U

表 2-2 HP Integrity rx2600 服务器主要竞争对手简表						
	IBM	Sun	Dell	IBM	Newisys	Dell
服务器	pSeries 615	Sun Fire V280,V480	PowerEdge 5250	xSeries 382	2100	PowerEdge 4600

处理器	Power4+ 1.45GHz	UltraSparc III 1.2 GHz	Itanium2, 1.5GHz	Itanium2, 1.5GHz	Opteron 242	Xeon 1.8-3 GHz
个数	1-2	1-2, 1-4	1-2	1-2	2	2

表 2-3 HP Integrity rx5670/rx4640 服务器主要竞争对手简表

	IBM	Sun	Bull	IBM	Newisys	Dell
服务器	pSeries 630	Sun Fire V880, V2400	NovaScale 4040	xSeries 450	4300	PowerEdge 6650
处理器	Power4+ 1.45 GHz	UltraSparc III 1.2 GHz	Itanium 2, 1.5GHz	Itanium 2, 1.5GHz	Opteron 844	Xeon MP 2.5, 2.8GHz
个数	1-4	1-8, 1-12	1-4	1-4	4	4

表 2-4 入口级服务器市场竞争总体势态

	与 64 位 RISC 的竞争	与 IA-32 的竞争	与 Opteron 的竞争
IPF 的优势	基于工业标准体系结构；提供领先的性能、性价比和较低的总拥有成本；	具有 64 位寻址能力；提供更高的性能；与中高档服务器的兼容性；	性能优势，特别是浮点性能是 Opteron 的一倍以上；拥有更多的 ISV 支持和解决方案，技术上也渐趋成熟；Intel 比 AMD 具有更强的实力和市场份额；
其他架构处理器的优势	拥有较多的解决方案、应用软件和成功实例；仍然占领主要高端市场，对入口级市场有推动作用；	大批量和普及应用；极其丰富的解决方案和应用软件；	价格低廉，高性价比；提供 64 位寻址能力和与 IA-32 的兼容性；价格比较便宜；
发展趋势	IPF 相对于 64 位 RISC 性能的优势将进一步扩大，批量的增大也将进一步扩大 IPF 性价比优势和市场份额；随着 IPF 平台上解决方案和应用软件增加、IPF 在企业级应用优势的确立，基于 IPF 入口级服务器将具有更大的竞争优势；	虽然 IPF 在批量和性价比方面很难与 IA-32 竞争，但是随着多媒体、数字成像、高端的游戏等应用的普及，电信和网络服务供应商规模的扩大，基于 IPF 的入口级服务器将具有更大市场前景；	Opteron 目前还与 Itanium2 不属于一个竞争档次；Opteron 是否能够得到足够的 ISV 支持、解决方案数量是否能够快速增长、所期望的 64 位桌面应用是否成熟都是一个未知数；

表 2-5 HP 的 zx1 与 Intel E8870 以及 IBM 的 Summit 芯片组的比较。

	HP zx1 芯片组	Intel E8870 芯片组	IBM Summit 芯片组
适用范围	经济实惠地适用于入口级服务器和工作站	主要应用于入口级服务器，芯片个数比 zx1 多。	主要应用于支持中档服务器，考虑了支持 Xeon

		可扩展到支持中档服务器	和 Itanium2 种处理器
处理器个数	1-4 个 Itanium2	1-4 个 Itanium2 在增加了可选的扩展端口后, 可支持 8-16 个 Itanium2 处理器	4-16 个 Xeon 或 Itanium2
CPU 带宽	6.4 GB/s	6.4 GB/s	6.4 GB/s
内存带宽	最大 12.8 GB/s	6.4 GB/s	6.4 GB/s(在 450 上 9.2)
内存延迟 (传输速度)	低 (快)	由于从 RD 转换成 DDR 延迟较高 (速度较低)	高(低)比 zx1 慢 4-9 倍
I/O 带宽	在 PCI-X 上 4 GB/s	在 PCI-X 上 4 GB/s	在 PCI-X 上 4 GB/s
最大内存 (使 用 2GB DIMM)	128 GB 256 GB/4GB DIMM	128 GB	64 GB (今天的 450 上 只有 40GB)
其他	内置的 AGP 图形功 能可选的内存缓存	支持 Snoop 模式的非均 匀内存访问	支持 cc-NUMA 非均匀 内存访问

表 2-6 HP zx1 芯片组相对于 IBM Summit 芯片组的优势

HP zx1	IBM Summit	特性	对用户的好处
Yes		高内存带宽、低内存延迟	最高的应用性能, 快速执行解决方案
Yes	Yes	大内存容量	能够访问内存中大数据集, 减少磁盘交换, 快速执行解决方案
Yes		支持 AGP-4X	在工作站上提供高性能 3D 图形功能
Yes		优化的 1-4 路可伸缩性	提供性能优化的 1-4 路基于安腾系统满 足各种需要
	Yes	优化的 4 路以上可伸缩性	支持更大的系统, 但将增加系统延迟不利 于支持 4 路批量较大系统
	Yes	内存镜像	增加内存可靠性, 但对大批量的中低档系 统好处不大, 且将增加内存成本
Yes		低成本	每单元 IT 投资提供更高的性能

表 2-7 rx2600 是多种操作系统下性能最高的 2 处理器服务器

测试指标	系统	操作系统	指标	排名
SPECfp_2000 rate	rx2600	Linux	42.4	2 路 #1
SPECint_2000 rate	rx2600	HP-UX	30.5	2 路 #1
TPC-C	rx2600	Windows	60,121	2 路 #1
Linpack N=1000 N*N N*N	rx2600	HP-UX	5,303 MFLOPS 9,853 MFLOPS	超过 IBM 和 Sun 最佳的 1 和 2 路
SPEC jbb 2000-java	rx2600	HP-UX	60,225	2 路 #1
SPEC web SSL	rx2600	HP-UX	1,930	2 路 #1

表 2-8 rx5670 是多种操作系统下性能最高的 4 处理器服务器

测试指标	系统	操作系统	指标	排名
SPECfp_rate (base)	rx5670	Linux	66.4 4-way 42.6 2-way	超过 IBM 和 Sun 最佳的 2 和 4 路
SPECfp_rate (base)	rx5670	HP-UX	59.8 4-way 30.3 2 way	4-路 #1
TPC-C	rx5670	HP-UX/Oracle	131,639@\$7.25/tpmC	4-路/UNIX #1
TPC-C	rx5670	Windows/SQL	121,065@\$4.79/tpmC	4-路/Windows #1
SPEC 2000-jbb	rx5670	HP-UX	116,466	4-路 #1
SPEC web SSL	rx5670	HP-UX	3,702	4-路 #1
TPC-C	rx5670	HP-UX/Oracle	131,639@\$7.25/tpmC	
TPC-C	rx5670	Windows/SQL	121,065@\$4.79/tpmC	4-路/Windows #1
SAP SD 2-层	rx5670	HP-UX/Oracle	860	4-路 #1
Linpack N=1000 N*N N*N N*N	rx5670	HP-UX 1 CPU 2 CPU 4 CPU	5,683 MFLOPS 11,490 MFLOPS 21,713 MFLOPS	超过 IBM 和 Sun 最佳的 1,2,4 路测试指 标
Oracle Apps	rx5670	HP-UX Linux	6,440 5,992	4 路 #1

### 领先的 SPEC Rate 基准测试指标

SPECfp\_rate2000 和 SPECint\_rate2000 是测试计算机系统多处理器浮点和整数计算能力的重要基准测试指标。入口级服务器一般装备 2-4 个处理器，因此 SPEC Rate 基准测试指标成为考察入口级服务器在设计和实施上是否能够充分发挥多处理器潜力、提供超过单处理器桌面系统性能的重要量度指标。由图 2-4 中的数据可见，基于 Itanium2 的 Integrity 系列入口级服务器提供超过装备 64 位 RISC、IA-32 和 Opteron 处理器的服务器一倍到二倍多处理器浮点和整数计算能力，使之更加适合于各种计算密集的应用。

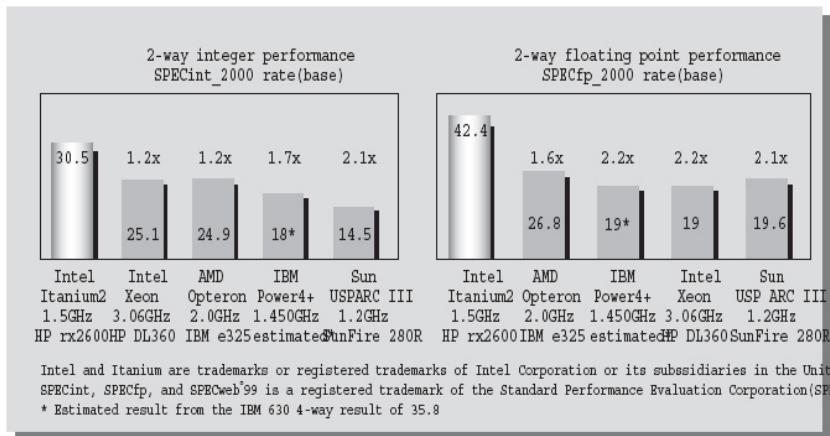


图 2-4 Integrity 入口级 rx2600 服务器提供领先的多处理器性能

## 领先的 tpmC 指标

TPC-C 测试给出系统每分钟进行交易数量 tpmC 指标，是企业用户测量系统在线事务处理(OLTP)和支持同时用户数能力最常用的指标。基于 4 个 1.5 GHz Itanium2 的 rx5670 服务器创造了高达 121,065 的 tpmC 指标(在 Windows/SQL 下)，不仅超过目前所有 4 处理器服务器，而且超过 IBM 和 Unisys 8 处理器服务器的最高指标的 1.2 倍以及 Sun Fire V 系列的 12 路服务器。装备 2 个 1.5 GHz Itanium2 处理器的 rx2600 的 tpmC 指标达到 60121，不仅高于装备任何体系结构处理器的 2 路服务器，而且高于 SunFire V 系列的 4 路服务器。图 2-5 和图 2-6 说明 HP Integrity 入口级服务器提供领先的 OLTP 性能。

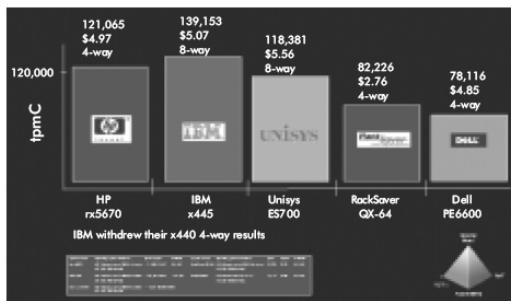


图 2-5 4-路 HP Integrity 服务器性能超过其他厂商 8-路服务器，价格更加便宜

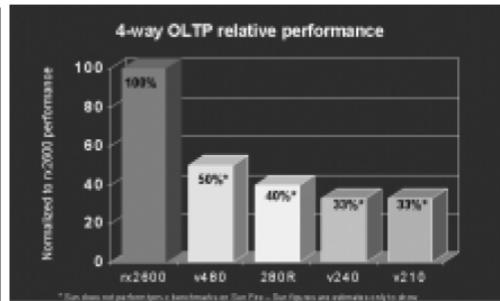


图 2-6 rx2600 性能超过所有 Sun V 系列所有 2 路服务器

## 领先的 SPECweb99 SSL 联接指标

SPECweb99-SSL 基准测试测量服务器有效地处理安全加密的 Web 交易的能力，是企业用户在 Internet 上执行高度安全的交易所需的关键指标。目前入口级服务器被大量应用与网络边缘，作为 Web 服务器或防火墙，处理安全加密的 Web 交易的能力对于考察入口级服务器的性能具有特别重要的意义。HP 与领先的 Web 服务器基础设施供应商 Zeus Technology 公司合作，在 rx5670 上运行 Linux 下的 Zeus Web 服务器软件创造了业界领先的 4 处理器系统，实现了支持 3702 个同时联接。该指标稍高于基于 IBM 基于 1.7GHz Power4+ 处理器的 8 路 p655(实际配置 4 路)的指标 3699，比基于 AMD 2GHz Opteron 的 4 路服务器指

标高 10%，是 Sun 基于 900-MHz UltraSPARC III 的 V480 的 6.5 倍以上。基于 1.5 GHz Itanium2 的 rx2600 服务器的 SPECweb99\_SSL 基准测试指标达到 1930，是 Sun 的 2 路 V240 和 V280R 相应指标只有 833 和 1008，只相当于 rx2600 50% 左右。

### 领先的 Oracle Apps 应用服务器指标

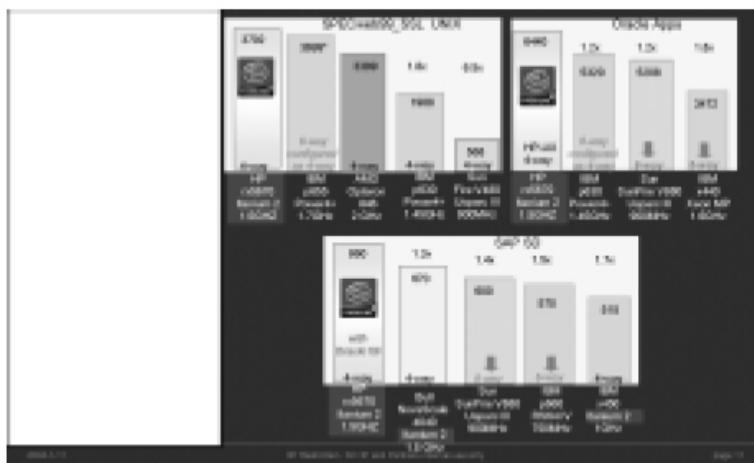


图 2-7 HP 基于 Itanium2 的 Integrity rx56704-路服务器应用性能超过 Sun 和 IBM 更大系统

Oracle 应用服务器是应用最广的应用服务器软件，为用户开发和运行企业应用软件提供可靠和完善的平台。图 2-7 表示 rx5670 提供领先的 Oracle Apps 指标，非常适合作为安装和运行 Oracle 应用服务器的平台。

### 领先的 SAP 基准测试指标

SAP 是领先的 ERP 软件厂商，SAP R/3 是应用最广的企业应用软件。为了测试各个厂商服务器产品的相对性能，SAP 设计了一系列基准测试。例如：SAP 销售和发货基准测试指标 SAP SD 和 SAP 需求分析和优化规划基准测试指标 SAP APO-DP 是测量服务器运行 SAP 应用软件完成供应链需求规划等企业应用能力的重要指标，体现了服务器支持企业应用的能力。HP 与 SAP 有长期的合作历史，双方在推动基于 IPF 服务器企业应用方面也进行了紧密的合作。HP Integrity 4 路 rx5670 服务器提供领先的 SAP 基准测试指标，非常适合于满足中小企业管理和大型企业工作组应用。rx5670 已经完成了第一个得到 SAP 正式认证 Windows 下 SAP APO-DP 基准测试指标，每小时完成 157,555 次规划组合操作，超过过去公布的 4 路系统最佳指标 15%，超过过去公布的 8 路系统最佳指标 21%。图 2-7 表示 rx5670 还提供领先的 SAP SD 基准测试指标。

### 领先的 SPECjbb2000 基准测试指标

SPECjbb2000 基准测试指标用来测试服务器一侧的 Java 性能，提供测量服务器运行

J2EE (Java2 企业版)能力最客观和代表性的基准测试指标。运行 HP-UX Java 的 rx5670 服务器实现了业界最佳的 4 处理器 SPECjbb2000 基准测试指标，每秒完成 116,466 次操作。这一性能水平远高于 IBM 公布的 4 处理器 RISC 系统最佳指标和 Sun 公布的 8 处理器 RISC 系统最佳指标高。基于 1.5 GHz Itanium2 的 rx2600 服务器实现了业界最佳的 2 处理器 SPECjbb2000 基准测试指标，每秒完成 60,225 次操作，比基于 1.3 GHz Power 4 的 IBM 工作站快 50%，比基于 1.05 GHz UltraSPARC III 的 Sun 工作站快 180%。

## 2.4.4 实现最佳的价格/性能

HP 在系统设计技术、支持多操作系统平台、连续可用性、高可伸缩性和可管理性等方面的优势，大大降低了它基于安腾产品系列管理、维护和升级费用、空间占用量、电源消耗和故障损失，使之具有最佳的性价比和最低的总拥有成本。

### 最高的性能密度

计算机系统的性能密度描述达到规定性能占用的空间量。提高性能密度不仅能够缩小机房面积，而且能够降低能耗、方便管理，许多厂商都力图借此降低系统的总拥有成本。当前许多电信企业和网络服务器供应商采用大量低端服务器提供服务，要求服务器占用尽可能少的空间。因此，性能密度成为入口级服务器的一个重要指标，对于降低用户的总拥有成本具有重要的作用。HP 的优化设计技术使它的基于 Itanium2 的入口级服务器结构十分紧凑、提供远比竞争对手高的性能密度。特别是，HP 推出了高度只有 4U 的 4 路服务器 rx4640，进一步提高了性能密度的优势。图 2-8 和图 2-9 分别说明 rx4640 和 rx2600 提供同档次系统中最高的性能密度。

HP 即将推出高度只有 1U、支持 2 路 Itanium2 的 cx2600 服务器，必将进一步提高 HP Integrity 系列入口级服务器的性能密度优势。



图 2-8 HP Integrity rx2600 超过 IBM 和 Dell 最佳的机架优化性能

图 2-9 HP Integrity rx4640 提供本档次最佳的性能密度

### 最佳的性价比

HP 基于 Itanium2 的 Integrity 系列在许多重要的应用中都提供优于竞争对手的性价比。随着 Itanium 技术发展和批量的增大、HP 市场的扩展，这一优势将进一步扩大。

### 最低的总拥有成本

随着企业间竞争的加剧，计算机系统的总拥有成本受到越来越多的重视。基于 Itanium 产品系列借助于 HP 的优势、提供最低的总拥有成本，使它们受到企业用户的广泛欢迎。

#### 2.4.5 对基于 Opteron 处理器服务器的竞争优势

2003 年 4 月 AMD 宣布推出基于 x86-64 (AMD64) 架构的 64 位 Opteron 处理器。AMD 声称 Opteron 突破了 32 位处理器 4GB 寻址空间的限制、与 IA-32 二进制兼容、能够同时执行 32 位和 64 位应用、提供最佳性价比和向 64 位过渡的平滑途径，一时间似乎出现了“另一个工业标准 64 位处理器”。虽然 Opteron 受到媒体关注和少数厂商支持，但事实上 Opteron 基于在原有的 x86-32 基础上改进的架构与基于革命性的 EPIC 架构的 Itanium2 不是一个档次的处理器，无论从所基于的处理器、系统技术、产品性能和厂商支持来分析，HP Integrity 系列入口级服务器与一些厂商基于 Opteron 的服务器相比较都具有全面的竞争优势。

### 处理器特性和性能

从基础架构来分析，Intel Itanium 处理器基于革命性的 EPIC 架构，而 Opteron 基于在成熟的 x86-32 基础上扩展的 x86-64 架构，使得 Itanium2 的并行特性、性能、可伸缩性、发展前景都超过 Opteron。Opteron 实际上是一个 32 位和 64 位混合的处理器，而 Itanium2 提供完全的 64 位计算能力、且与 IA-32 完全二进制兼容。

IPF 许多领先的特性使得 Itanium2 提供远比 Opteron 高的基准测试指标。

### 市场上竞争性的基于 Opteron 服务器产品

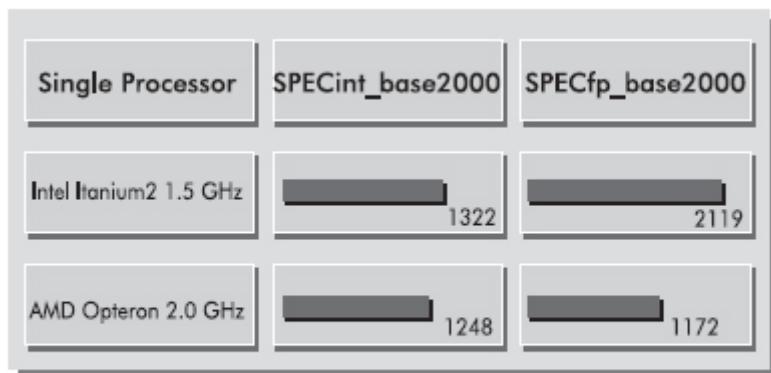


图 2-10 Itanium2 领先于 Opteron 的基准测试指标

由于 Opteron 上市不久、只受到有限的 IHV 的支持、加上本身可伸缩性的限制，目前市场上只有与 Integrity 入口级服务器相对应的 2-4 路的 Opteron 服务器。具体如表 2-12 和图 2-11 所示。

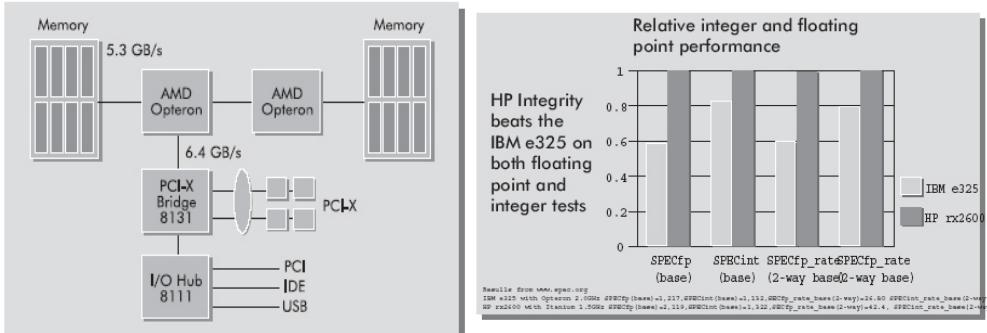


图 2-11 基于 Opteron 处理器的 2 路服务器方框图

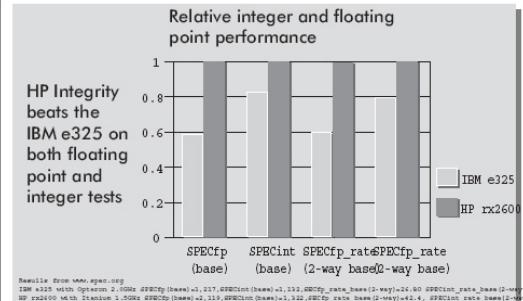


图 2-12 HP Integrity rx2600 提供领先于 IBM e325 的整数和浮点性能

Integrity 入口级服务器提供相对于 Opteron 2 路和 4 路服务器更加丰富的资源，满足应用的需求。

由表 2-13 可见，rx2600 不仅提供高于 Newisys 2100 的性能，而且提供 150% 内存容量、2 倍以上的 I/O 带宽，130% 的内部磁盘容量。2004 年将推出的 c2600 以 1U 高度、更低的价位满足电信等领域应用需求。

表 2-14 可见，rx5760 不仅提供高于 Newisys 4300 的性能，而且提供 3 倍最大内存容量、2.5 倍以上的 I/O 带宽。rx4640 以同样的高度提供更高的资源容量和扩展空间。

## HP 领先的芯片组技术

HP Integrity 系列入口级服务器基于 HP zx1 芯片组与一些厂商使用的 AMD 8131 芯片组比较具有一系列领先的特性。(见表 2-15)

### HP Integrity 系列入口级服务器的性能优势

Itanium2 领先的性能和 HP zx1 芯片组领先的特性，使得 HP Integrity 系列入口级服务器提供领先于其他厂商基于 Opteron 的竞争性服务器的性能和可伸缩性。

图 2-12 表示 HP Integrity rx2600 性能领先于 IBM e325。AMD 虽然声称 Opteron 提供更高的扩展线性。事实上，HP 领先的芯片组和系统技术使得 HP Integrity 系列入口级服务器具有领先的可伸缩性(图 2-13)。

### 生态系统优势

从系统建设角度来分析，HP Integrity 服务器更是大大领先于一些厂商的 Opteron 服务器。HP 与许多领先厂商合作，在 IPF 平台上进行了全面的系统建设，建立了完整的生态系统即 Integrity 系列，为广大用户和 ISV 提供从操作系统、编译程序、工具软件、数据库系统到系统基础件(中间件)、业务基础件、完全可以与 64 位 RISC 相比美多的完整应用开发平台(详见[14])。HP 也在 IPF 平台上建立了支持企业关键任务运行的环境(详见[14])。对比之下，Opteron 服务器的系统建设还刚刚开始，要赶上 HP 的 Integrity 系列还需要很长的时间。

## 厂商支持优势

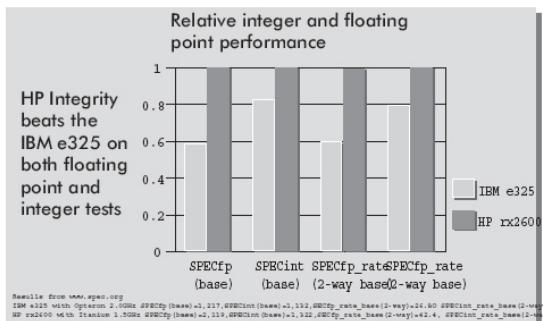


图 2-12 HP Integrity rx2600 提供领先于 IBM e325 的整数和浮点性能

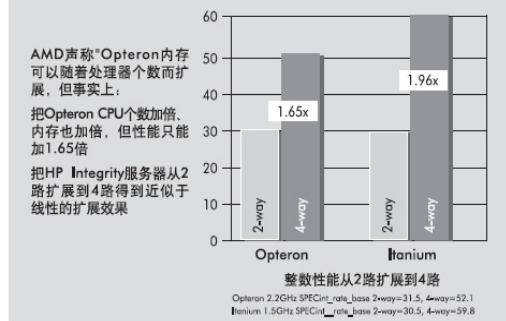


图 2-13 基于 HP Itanium 的系统提供超过基于 Opteron 系统的扩展线性

IPF 已经获得了 40 多个 IHV 的支持，提供 8 路以上服务器产品的 IHV 也已经超过 10 家。目前已经有从入口级到企业级服务器进入市场，而 Opteron 只得到了不多几家 IHV 的支持，产品也仅局限于入口级档次。IBM 虽然也推出了一款 2 路 Opteron 服务器。实际上，IBM 大力支持的是本公司专属的 Power 处理器，IBM 也支持 Itanium，最近推出了 16 路的 Itanium2 服务器。IBM 对 Opteron 的支持只是象征性的，或者仅仅局限于利用基于 Opteron 的低端系统构建集群。IPF 已经获得了 400 多家 ISV 的支持包括所有一流的企业应用和 HPTC 领域的 ISV，IPF 平台上应用软件的数量已经接近 1000 个。预计到 2004 年中，IPF 平台上 ISV 和应用软件将分别超过 1000 个和 2000 个(详见[14])。相反，许多 ISV 对 Opteron 还处于评估阶段，是否在这一平台上真正推出优质的软件产品还是一个未知数。

总之，Opteron 服务器与 HP Integrity 系列入口级服务器不是一个档次的产品。后者不仅目前全面领先于 Opteron 服务器，而且拥有更加远大的发展前景。

## 三、HP Integrity 系列中档服务器

中档服务器的一般特征是：装备 8-16 个 CPU、价格不超过 100 万美元、在中小企业数据中心或大企业部门层次应用。HP 基于 Itanium2 的 Integrity 系列目前包括 rx7620 和 rx8620 等两个中档服务器。HP 通过全面贯彻其在工业标准部件基础上进行增值的战略，使 HP 基于工业标准 Itanium2 的中档服务器提供最佳的价格/性能、实现“以中档服务器的价位提供主机(或企业级服务器)的性能”，并具有最高的兼容性和可靠性、最广泛的操作系统选择余地和最全面的解决方案，提供领先于其他厂商基于 Itanium2 和 RISC 处理器中档服务器的竞争优势。本章介绍 HP Integrity 系列中档服务器的产品概貌、体系结构、特性和竞争优势。

### 3.1 HP Integrity 系列中档服务器概述

HP Integrity 系列提供 rx7620 和 rx8620 等型号的中档服务器，其硬件的基本参数如表 3-1 所示。

### 3.2 HP Integrity 系列中档服务器架构

HP Integrity 系列中档服务器采用 cc-NUMA 架构、分层交叉交换互联拓扑。这一架构是基于 HP 专门设计的 sx1000 芯片组实施的，奠定了全面领先优势的基础。

#### 3.2.1 HP Integrity 中档服务器架构

rx7620 架构是围绕把系统作为一个 2-8 路 SMP 系统运行或划分成 2 个独立硬件分区 (nPars) 的要求设计的。图 3-1 表示 rx7620 架构的主要部件。当该系统配置成非分区服务器时，所有资源作为一个逻辑服务器执行。当它配置成两个 nPars 时，系统资源被划分成两个逻辑服务器或独立分区，每个分区包含一块拥有自己专用 I/O 资源的单元板。假如在图 3-1 中联接上下单元板的实线不复存在，该图将表示一个划分成两个独立分区的系统。上半图中的单元板、I/O 港湾、核心 I/O 和外设港湾将变成一个独立的硬件分区，它将与下半图中的第二个分区完全隔离开来。

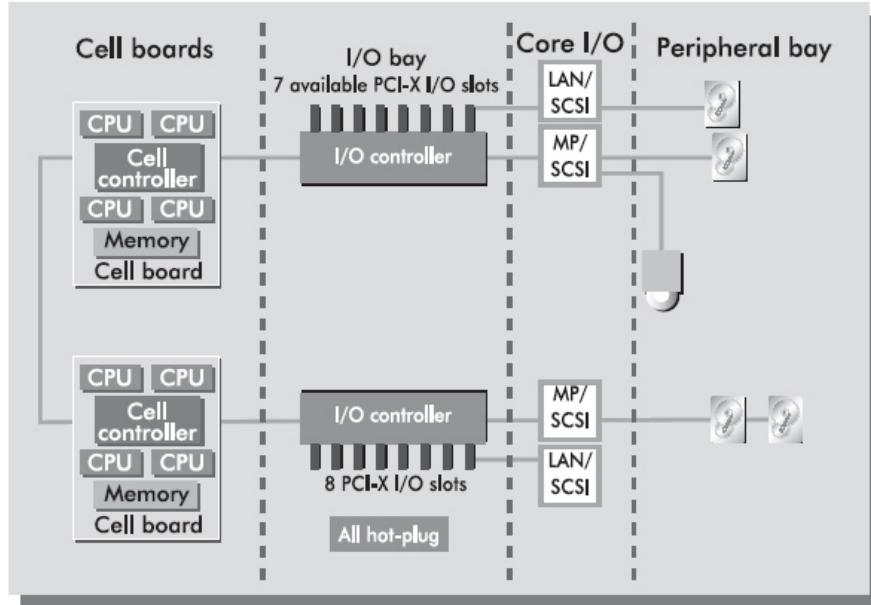


图 3-1 HP Integrity 服务器 rx7620 架构

rx8620 架构建立在 rx7620 架构以及一个交叉交换后面板和两个附加的单元板上。交叉交换后面板提供最多 4 个单元间的非互锁联接、再加上到 HP Server Expansion Unit (SEU) 中外部 I/O 资源间的联接。与 rx7620 相类似，rx8620 可以配置成一个 2- to 16-way SMP 系统，它也可以划分成较小的独立 nPars。当联接到 SEU 时，rx8620 可划分成 4 个硬件隔离的分区。

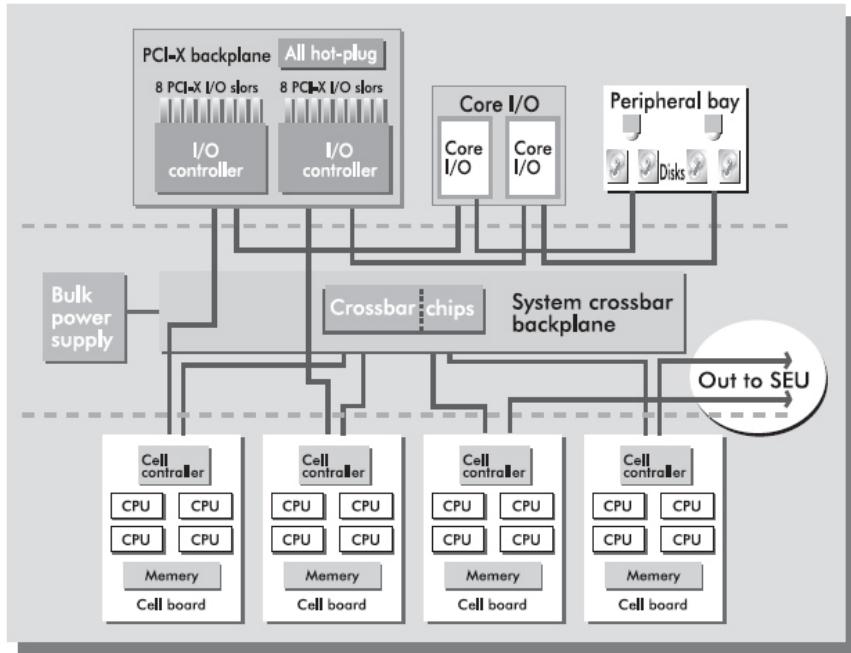


图 3-2 HP Integrity rx8620 服务器架构

### 3.2.2 单元板

单元板是 Integrity 中档服务器的基本组成构件。每个单元板都是一个独立的部件，包括：对称多处理器 (SMP)、主内存以及所有必需的硬件：

- 最多 4 个处理器模块；
- 单元控制器 ASIC(专用集成电路)；
- 主内存 DIMM，最大容量为 32GB(每个板最多包含 16 个 DIMM，4 个 DIMM 为一增量，使用 512 MB 或 1 GB DIMM 或者两者的组合)；
- 稳压器模块(VRM)；

#### 单元控制器

单元控制器 ASIC (CC) 处于单元板的心脏。CC 提供处理器、内存、I/O、处理器相关硬件和相邻单元之间的通信路径。单元控制器芯片包含接口逻辑和负责维持整个系统缓存一致性。与单元控制器 ASIC 相联接的是最多 4 个 Intel Itanium 2 处理器和最多 32 GB 主存。每个单元直接或者通过交叉交换器后面板与相邻的单元以及 I/O 资源相联接。

#### 内存控制器 ASIC

内存控制器 ASIC 也是 sx1000 系统芯片组的组成部分，它的主要功能是多路传输和多路分解。单元控制器 ASIC 与内存子系统中 SDRAM 之间的数据。当单元控制器 ASIC 在内存接口命令总线上发布一个只读操作时，内存控制器 ASIC 缓冲 DRAM 读取数据，并尽快送回。当单元控制器 ASIC 发布写操作时，内存控制器 ASIC 接收来自单元控制器 ASIC 的写数据，

并将其转给 DRAM。

注意，只有内存子系统的数据部分通过内存控制器 ASIC，所有的 DIMM 地址和控制信号都由单元控制器 ASIC 产生，然后通过内存接口地址总线直接发送给 DIMM，从而缩短了内存的延迟时间。

内存子系统具有四通路的通信机制，支持内存 DRAM 容错，即使一个独立的 SDRAM 芯片发生故障，也不会影响数据的完整性。内存子系统为单元控制器 ASIC 提供 16 GB/s 的峰值带宽，并将典型情况下与目录一致性相关的内务操作最小化，不仅如此，内存子系统从单元到本地内存访问的延迟时间也很短：加载使用时的平均空闲延迟时间只有 245 ns。

## 单元配置

rx7620 支持最少 1 个、至多 2 个单元。rx8620 支持最少 1 个、至多 4 个单元。每个单元可以在购买 2 个或 4 个活动的 Itanium 2 处理器，或者活动处理器与按需供应立即容量 (iCOD) 处理器的组合。在单元内，CPU 到 CC 的峰值带宽为 12.8 GB/s，CC 到内存的峰值带宽为 16 GB/s。CPU 可以通过 CC 直接访问内存，因此不管有多少 CPU 都可以直接访问所有内存模块。

### 3.2.3 交叉交换背板

下一个基本构件是交叉交换背板。交叉交换背板包含两个交叉交换芯片，提供 4 个单元以及它们的内存和 I/O 之间的非互锁联接。(rx7620 不需要使用交叉交换后面板，因此它的单元间的通信是在一条直接联接总线上进行的)

## 交叉交换器 ASIC

交叉交换器 ASIC 是 sx1000 芯片组的另外一个组成部分，它实施高性能 8 路无阻断交叉交换通信机制和 500 MHz 交叉交换链路协议，所有端口在功能和电规格上都完全相同。rx8620 服务器拥有一个充分连接、无阻塞交叉交换网络，共有 2 对交叉交换器 ASIC。

交叉交换网格的一个非常重要的特性，是所有的链路具有相同的带宽和延迟时间。这对于最大限度地提高系统总体聚合带宽和最大限度地减少系统总体延迟时间具有十分重要的意义。单元到交叉交换器和交叉交换器到交叉交换器的通讯以相同的速度进行，从而控制访问远程内存的延迟时间，缩小访问本地和远程内存的延迟比。此外，rx8620 服务器的内存首先在单元之间、然后在内存条之间交错，这种交错设计可以平衡所有链路之间的内存通信量。

rx8620 的全程交叉交换器网络实现了一个全局的点到点包过滤网络，这个网状结构具有极高的完整性，每一个交叉交换通路完全独立。全程交叉交换器网络具有专用的数据和控制路径，每个通路可以完全独立于其它通路进行复位、分配或重新配置。rx8620 服务器的这一设计为资源隔离奠定了良好的基础。

交叉交换器 ASIC 提供一系列有助于提高 rx8620 服务器高性能的特性：

- 支持扩展到 128 路一致共享内存系统(采用 PA-8800、PA-8900 和 mx2 处理器)；
- 250 MHz 运行速度；
- 500 兆次传输/秒 (MT/s) 的链路速度；

- 链路协议支持两个交错式通道；
- 支持 Intel Itanium 处理器家族的双倍长度数据包模式；
- 性能计数器便于软件优化；

### 3.2.4 I/O 子系统

HP Integrity 系列中档服务器具有高性能、可伸缩、灵活和可靠的 I/O 子系统。

#### I/O 子系统

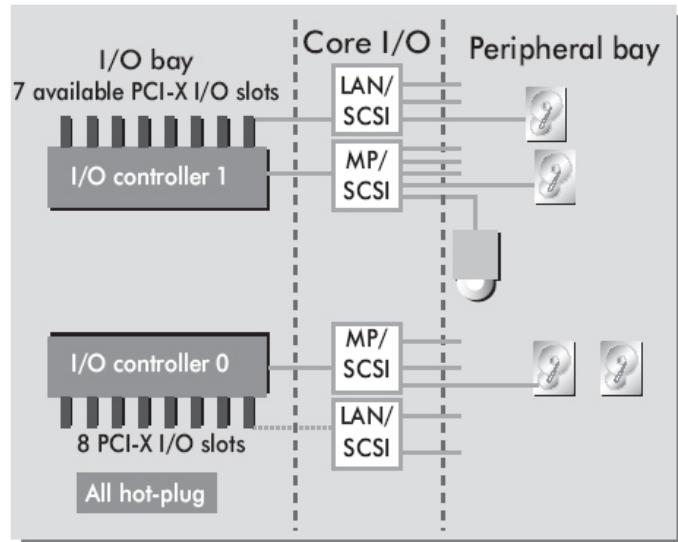


图 3-3 HP Integrity rx7620 服务器 I/O 子系统

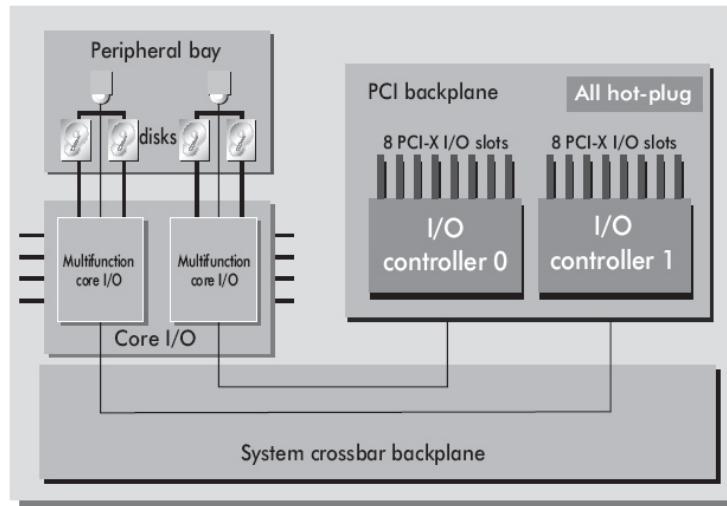


图 3-4 HP Integrity rx8620 服务器 I/O 子系统

每个 rx7620 和 rx8620 服务器包含一个嵌入高性能 I/O 子系统。此外，rx8620 可以根据用户选择通过一条高性能 I/O 电缆联线与放置在 HP Server Expansion Unit (SEU) 中的外部 I/O 资源相联接。I/O 子系统中的部件是 I/O 控制器、内部外设港湾和多功能核心 I/O。图 3-3 和图 3-4 表示 rx7620 和 rx8620 I/O 子系统的基本方框图。

#### I/O 控制器芯片

rx7620 和 rx8620 服务器包含两个主 I/O 控制器芯片放置在 PCI-X 后面板上。每个 I/O 控制器包含 16 条高性能 12 位宽的联线。这些联线联接到 16 个支持 PCI-X 卡插槽和核心 I/O 的从属 I/O 控制器芯片。

在两个系统中，两条联线(每个主控制器引出一条)专门用于核心 I/O。其余 30 条联线在 16 个 133 MHz x 64-bit PCI-X 卡插槽间划分，每个插槽都有一条专门的 PCI-X 总线。这种每条总线一块卡的设计提供更高的 I/O 性能、较好的容错能力和更高的可用性。

每个控制器芯片直接联接到一个主机单元板。这意味着为了访问所有 I/O 卡插槽必须购买两个单元板(一块单元板只能访问一半插槽)。

## PCI-X 后面板

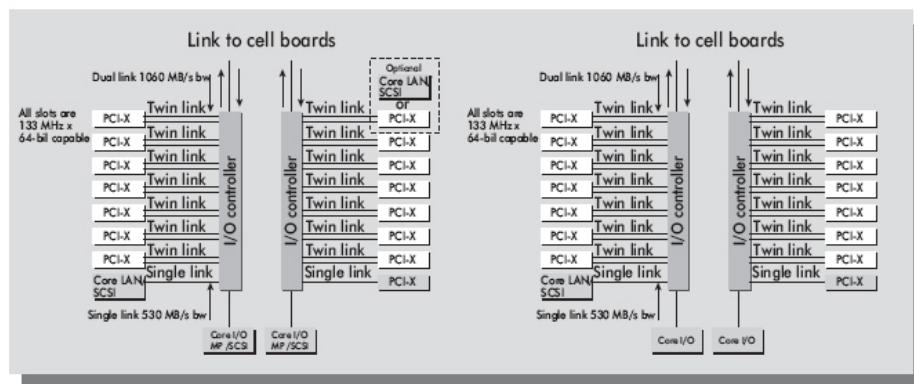


图 3-5 HP Integrity 系列 7620 和 8620 服务器后面板方框图

图 3-5 表示 HP Integrity rx7620 and rx8620 服务器 PCI-X 后面板的细图。两个服务器 I/O 插槽的实施几乎是完全相同的，不同之处是 rx7620 核心 I/O 使用 1-2 个插槽。在该图中，16 个 I/O 卡插槽中有 14 个受到双倍的高性能联线支持。这些双联线 I/O 插槽提供每个插槽最大为 1.06 GB/s 的峰值带宽。其余两个 I/O 插槽是单联线的，提供最大 530 MB/s 的峰值带宽。聚合 I/O 插槽带宽为 15.9 GB/s。rx7620 和 rx8620 服务器中所有 PCI-X 插槽能够以 133 MHz x 64 位速率运行这意味着每个 I/O 插槽都允许业界性能最高的 PCI-X 以最大的设计速度运行。

在实践中，需要最大带宽的 PCI-X I/O 卡应当插入双联线插槽。由于每个 I/O 插槽都有专门的总线，插槽都可以热插拔或者维修、而不会影响其他插槽。

## 核心 I/O

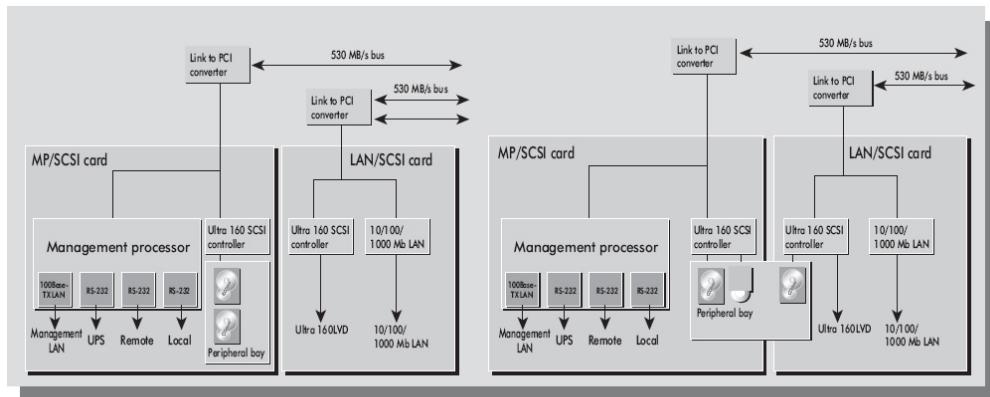


图 3-6 HP Integrity 系列 7620 和 8620 服务器核心 I/O 方框图

rx7620 和 rx8620 服务器可以按照具有 1 或 2 个核心 I/O 产品的系统来配置。在两个系统中，核心 I/O 提供控制台、Ultra160 SCSI、Gigabit LAN、serial 和管理处理器(MP)功能以及外设港湾的 SCSI 控制器。第二个核心 I/O 产品可应用于建立第二个分区，提供访问第二组磁盘驱动器的能力和提供冗余 MP 功能。在 HP Integrity rx8620 服务器中，第二个核心 I/O 产品也能够实现使用附加的可移动介质设备。

### 3.3 HP Integrity 中档服务器特性

HP Integrity 系列中档服务器采用与企业级 Superdome 服务器相同的 cc-NUMA 架构和分层交叉交换互联架构，提供领先中档性能、领先的中档灵活性、领先的投资保护、领先的中档使用价值，实现了以中档价格、企业级性能和特性的设计目标，成为用户同时满足支持关键任务和高强度计算应用以及经费约束等两方面需求的最佳选择。表 3-2 列出 HP Integrity 系列中档服务器的一系列领先的特性(进一步细节见 4.3 节)。

### 2.4 HP Integrity 系列中档服务器竞争优势

中档服务器是服务器市场竞争的重要领域。HP 是这一领域领先的厂商，占有 33.7% 的市场份额。Itanium2 领先的性能和 HP 先进的系统技术奠定了 HP Integrity 系列中档服务器竞争优势的坚实基础。HP Integrity 中档服务器在价位和资源容量之间作出了最佳的选择，不仅具有在许多重要和代表性浮点计算(如 Linpack)和企业应用(如 SPECwebSSL, SPECjbb, TPC-C, SAP 和 Oracle 应用服务器等)的性能指标都领先于竞争对手，而且具有领先性价比和最低的总拥有成本，成为各种 HPTC 和企业应用的最佳选择。

#### 2.4.1 市场竞争态势

中档服务器在产品系列中起承上启下的关键作用：入口级服务器由于价位的限制、资源容量有限、性能也不可能太高，企业级服务器为了面向高端的应用需求必须配置很大的资源容量、价格一般比较昂贵。因此，中档服务器的竞争焦点是在性能和价格之间找到一个最佳的折衷点：提供最佳的价格/性能即实现所谓以中档服务器的价位提供企业级服务器的性能。当前，各厂商都力图提供兼有入口级和企业级服务器在价位和资源容量两方面优势的中档服务器，满足广阔的市场需求，使中档服务器成为又一重要的竞争领域。过去 Intel IA-32 体系结构处理器在服务器市场中份额很小，更难进入中档服务器的应用领域。Xeon 的成功使得 Intel 在服务器市场不仅在低端夺取了 RISC/UNIX 服务器相当大的市场份额，而且也成功地打入中档服务器市场。因此，HP Integrity 系列中档服务器在市场上不仅需要面对其他厂商基于 Itanium2 和 RISC 处理器的中档服务器的竞争，而且需要面对基于 Wintel 工业标准的 32

位 Xeon/Xeon MP 中档服务器的竞争。

## 2.4.2 工业标准平台上的技术优势

HP 在工业标准技术基础上进行创新和增值的产品总体战略，奠定了 HP Integrity 系列的技术优势。对中档服务器，这一战略体现在基于工业标准平台、在价格与系统资源和性能之间找到理想的折衷点，实现了以中档服务器的价格提供企业级的性能，以最佳的性价比满足企业要求的设计目标。HP 作为 IPF 的共同开发者，能够最透彻地了解硬件内部特性。HP 基于 IPF 的服务器在硬软件设计方面都能够充分发挥 Itanium2 处理器的性能优势和潜力：HP 领先的 sx1000 芯片组支持 HP Integrity 中档服务器配置最丰富的中档层次资源、拥有充分资源潜力的基础设施和领先的可伸缩性；在系统软件方面，HP Integrity 系列中档服务器的操作系统、编译程序和开发工具包等软件都针对处理器和芯片组的特点进行了优化，进一步增强了整个系统性能和性价比领先地位，使得 HP Integrity 系列中档服务器成为用户同时满足经费和性能两方面要求的最佳选择。

### 领先的 Itanium 2 服务器和 sx1000 芯片组

HP Integrity 系列中档服务器基于 Intel Itanium2 处理器和 HP sx1000 芯片组，它们采用当前最领先的芯片技术、奠定了全面领先于基于 RISC 和 IA-32 处理器的坚实的基础。从发展趋势看，Intel IPF 处理器系列性能提高的速度也将高于其他处理器系列，从而确保 HP 基于 IPF 的服务器产品的长期领先地位。（详见[13]）。

### 领先的中档系统资源

中档服务器可以拥有相当大的系统资源，其中有些已经能够提供接近企业级服务器的资源容量。HP Integrity 系列 rx7620 和 rx8620 提供本档次内领先的系统资源。

Integrity 中档服务器能够在允许的价格范围内，提供实现领先功能的系统资源配置和高性价比，包括：

- **较多的 CPU：** Itanium2 领先的性能，使装备 8, 16 个 Itanium2 的 Integrity 中档服务器能够提供非常高的信息处理能力；
- **较大的内存和磁盘容量：** Integrity 中档服务器能够提供 64-128 GB 的内存容量和 TB 级的磁盘容量，具备支持较大规模数据库和数据仓库应用的潜力；
- **较大的 I/O 和网络通信能力：** Integrity 中档服务器具有较强的 I/O 吞吐和网络通信能力以及各种标准的网络协议，具有作为相当规模企业的网络和存储服务器的能力；

### 充分发挥资源潜力的系统基础设施

Integrity 中档服务器采用 ccNUMA 体系结构、两层交叉交换互联拓扑和 HP 专利的 sx1000 芯片组建立了领先的系统基础设施，提供高带宽、低延迟、消除了系统瓶颈（详见第 4 节），同时保持低成本，使系统资源能够充分潜力，实现了中档服务器高性能和高性价比完

美结合的关键优势。HP Integrity 中档服务器借助 sx1000 的优势，能够提供更高的系统带宽和 I/O 带宽，为提供高性能、线性可伸缩性和支持企业应用能力奠定了坚实的基础。

### 高可伸缩性和投资保护

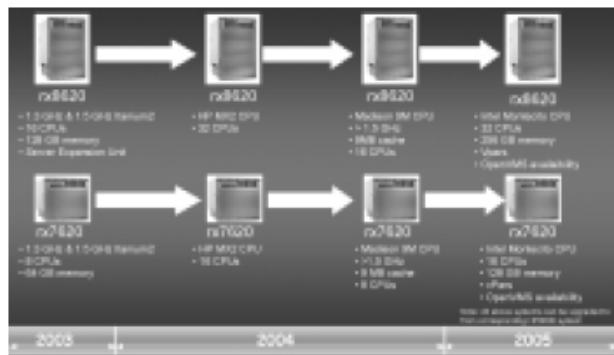


图 3-7 HP Integrity 系列中档服务器的可伸缩性

中档服务器系统资源较丰富、投资较大、软件积累较多、应用面和影响面也较广，因此中档服务器在扩展和升级时要更加强调投资保护和效率。HP Integrity 系列 rx7620 和 rx8620 服务器将在 2004 年通过采用 mx2 扩展模块支持 16 和 32 个处理器、在 2004 通过采用装备 9 MB L3 缓存和主频更高的 Itanium2、2005 年采用双核的 Montecito 处理器提供更高的处理能力、支持更大的内存容量，实现最高的本系列内的可伸缩性(也有人称为机箱内可扩展性)。

HP Integrity 系列中档服务器的机箱内升级的特性，还允许基于 PA-RISC 的 HP9000 系列的中档服务器在机箱内由使用 PA-RISC CPU 升级为使用 Itanium2、以后还可以沿着使用工业标准的 IPF 系列处理器的道路继续在机箱内升级，从而保护用户原有的投资。

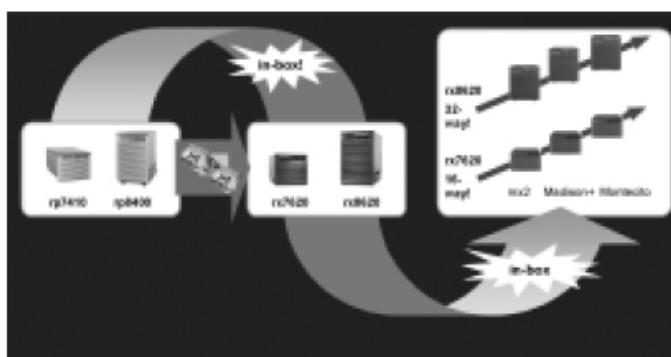


图 3-8 HP 9000 中档服务器可以在机箱内升级到 HP Integrity 系列中档服务器

### 2.4.3 提供领先的性能

HP 领先的系统技术和 Itanium2 的高性能相结合确保 HP Integrity 系列中档服务器具有本档次内全面领先于 RISC 处理器和 IA-32 处理器的基准测试指标和应用性能。

### 全面领先的基准测试指标和应用性能

HP Itanium 技术的优势确保其基于 Itanium2 中档服务器在主要的企业应用领域提供全

面领先于竞争对手同档次的系统，有力地促进 Itanium 向广泛的市场领域扩展。表 3-7 说明基于 1.5 GHz Itanium2 的 rx7620, rx8620 服务器已经成为多种操作系统下性能最高中档服务器，具有全面的领先地位。

## 领先的 tpmC 指标

TPC-C 测试给出系统每分钟进行交易数量 tpmC 指标，是企业用户测量系统在线事务处理(OLTP)和支持同时用户数能力最常用的指标。基于 8-16 个 1.5 GHz Itanium2 的 Integrity rx7620 和 rx8620 服务器的 tpmC 指标也居领先地位。例如，图 3-9 说明 HP 基于 Itanium2 的 rx8620 提供相当于 36 处理器服务器相当于 Sun Fire V 系列的 tpmC 指标。

## 领先的 SPEC Rate 基准测试指标

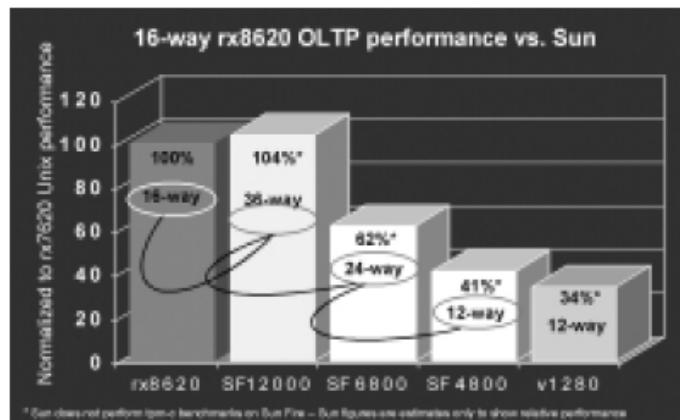


图 3-9 16-way HP Integrity rx8620 性能可以与 Sun 的 36- 路 Sun Fire 12K 竞争

SPECfp\_rate2000 和 SPECint\_rate2000 是测试计算机系统多处理器浮点和整数计算能力的重要基准测试指标。中档级服务器一般装备 8, 16 个处理器，因此 SPEC Rate 基准测试指标成为考察中档服务器在设计和实施上是否能够充分发挥多处理器潜力、提供较高的处理整数或浮点计算密集任务的能力。由表 3-9 中的数据可见，基于 Itanium2 的 Integrity 系列中档服务器提供超过装备 1 倍以上 64 位 RISC 和 IA-32 处理器的服务器浮点和整数计算能力，体现了 IPF 性能和 HP 领先技术的优势，也超过其他厂商基于 Itanium2 服务器的浮点和整数计算能力，充分体现了 HP 系统技术的优势。

表 2-10 HP 基于安 Itanium2 品具有最低的总拥有成本

	HP rx5670	Sun v880	IBM p660	Sun v480	IBM p630
购买成本	\$749060	\$1499970	\$2264832	\$1099945	\$604719
支持合同	\$56825	\$21184	\$178952	\$138864	\$113004
3 年机房费用	\$39600	\$118800	\$158400	\$79200	\$39600
3 年电费	\$18506	\$52488	\$46032	\$38277	\$26244
总费用	\$863991	1584642	\$2648216	\$1356286	\$783567

表 2-9 HP 基于安腾产品系列在 Web99 应用中提供最佳的性价比

	HP rx5670 Itanium 2 4-路服务器	Sun Fire v480 4 路 服务器	IBM p630 4 路服务器
每个服务器 价格	\$87000	\$50400	\$68200
SPECweb 99_SSL	3702	568	1050
处理每个交易能力的 价格	\$23.5	\$89	\$65

表 2-11 IPF EPIC 架构与 Opteron 的 x86-64 架构特性比较

	EPIC 架构	X86-64 架构	对性能的影响
并行机制	显性并行	隐性并行	显性并行利用编译程序来标识程序中可并行执行的指令段、允许使用硬软件的合力来提高处理器性能；隐性并行只能利用芯片逻辑来反复发现程序中可并行执行的指令段，不能利用硬软件的合力来提高性能；
寻址能力	64 位虚拟地址 50 位物理地址	48 位虚拟地址 40 位物理地址	Opteron 只能访问 1TB 的物理地址和 48 位虚拟地址，无完全 64 位寻址能力，今后向 64 扩展的工作量也较大
转移指令	利用硬软件合力优化转移指令的执行	只能利用硬件来预测转移地址	EPIC 能力通过编译程序和硬件逻辑的配合提供预测、暗示和预取等功能，实现零开支转移，消除转移指令的影响、提高取指令的缓存命中率，大大提高处理器的性能；x86-64 只能利用硬件预测转移地址、预测错误后开支很大，特别是 x86 架构的流水线很长（12 级），进一步增加了预测错误对性能的影响；
寄存器	拥有 384 个寄存器，包括 128 个整数和 128 个浮点寄存器	只有 40 个寄存器，其中 16 个整数和 16 个浮点寄存器	Opteron 处理器中寄存器数量太少，使之在应用于需要处理大量数据的 HPTC 和数据仓库等领域时，性能将大大下降。事实上，Opteron 在许多实际应用中与 Itanium2 的性能差距比基准测试指标的差距还要大
子程序调用	提供寄存器堆栈引擎（RTE）功能	不提供 RTE 功能	EPIC 提供的 RTE 功能能够通过改变堆栈指针、消除调用子程序时保存和恢复寄存器的开支。随着，应用软件开发向 Java 和 Web Services 方向发展，子程序调用频率越来越大，EPIC 的 RTE 优势也越来越大；
可伸缩性	目前能够支持 64 个处理器，	目前只有支持 4 个处理器的服务器	HP 已经推出基于 Itanium2 的全系列产品，全面支持入口级、中档和企业级应用（或网络边缘、应用服务器和数据库服务器三个层次的应用）；

	2004 年将支持 128 个处理器	产品, 最多能支持 8 个处理器	Opteron 目前只能支持入口级的服务器, 很难扩展到支持企业级应用

表 2-12 市场上基于 Opteron 的服务器产品

厂商	IBM	Newisys	Newisys
服务器	xSeries 325	Newisys 2100/Rack Server RSN 1164	Newisys 2100/Rack Server Quatrex-64
处理器	Model 240,242,246(1.4, 1.6, 2 GHz)	Model 242 1.6 GHz	Model 844
个数	2	1-2,1-4	4

表 2-13 Integrity 与 Opteron 2 路服务器比较

	HP rx2600	HP c2600 (将在 2004 年推出)	Newisys 2100	IBM xSeries 325
处理器	1-2 个 Itanium2 6M	1-2 个 低功耗 Itanium2	必须使用 2 个处理器 Opteron 242	Model 240,242,246 (1.4, 1.6, 2 GHz)
高度	2 U	1 U	1 U	1 U
最大内存	24 GB	512 MB-12 GB	16 GB (必须使 用 2 个 CPU)	1-6 GB
I/O 插槽	4 PCI-X	2 PCI-X	2 PCI-X	2 PCI-X
I/O 带宽	4 GB/s	2 GB/s	1.5 GB/s	1.5 GB/s
存储容量	3 个热插拔 SCSI 磁盘	2 个热插拔 SCSI 磁盘	2 个热插拔 SCSI 磁盘	36-292 GB

表 2-14 Integrity 与 Opteron 4 路服务器比较

	HP rx5670	HP rx4640	Newisys 4300
处理器	1-4 个 Itanium2 6M	1-4 个 Itanium2 6M	必须使用 4 个处理器 Opteron 844
高度	7 U	4 U	4 U
最大内存	96 GB	最大 64 GB	32 GB (必须使用 4

			个 CPU)
I/O 插槽	10 PCI-X	6 PCI-X	5 PCI-X
I/O 带宽	9.5 GB/s	4 GB/s	3.5 GB/s
存储容量	4 个热插拔 SCSI 磁盘	2 个热插拔 SCSI 磁盘	4 个热插拔 SCSI 磁盘

表 2-15 HP zx1 芯片组领先于 AMD8131 芯片组

	HP zx1 芯片组	AMD 8131 芯片组
适用范围	zx1 是入口级服务器和工作站，集成 AGP 支持	Opteron 包括集成的内存控制器。对入口级服务器和工作站，增加 AMD8151(PCI-X 加 AGP)或 AMD 8111(I/O 集线器 = “Super I/O”)
最大 CPU 数	1-4 Itanium 2	1-8 个处理器，使用 HyperTransport 互联技术
CPU 之间带宽	6.4 GB/s	6.4 GB/s
内存带宽	对 4-路系统最大 12.8 GB/s 对 1-2 路系统 8.6 GB/s	每个 CPU 5.3 GB/s. AMD 声称内存带宽随着 CPU 数量的增加而扩展，但是这将迫使 CPU 访问 cc-NUMA 架构中的远程内存，造成较大的延迟
内存延迟	很低 (快速)	访问本地内存延迟为 85n 访问远程内存，每次跨越延迟为 140n
最大内存容量	每个系统 128 GB (使用 2GB DIMM)，将来可达 256 GB (使用 4GB DIMM)	每个 CPU 最多 16 GB (8 个内存插槽，使用 2GB DIMM)，因此，只有 8 处理器系统才能达到 128 GB 内存容量
I/O 带宽	4 GB/s PCI-X，内置的 AGP	每个 AMD 8131 2GB/s PCI-X
RAS 特性	校正 4 位错误，探测 8 位 错误	校正 1 位错误，探测 2 位 错误
其他特性	内置的 AGP，支持利用内存缓存区扩展内存容量	必须使用不同的桥接芯片支持 AGP，每个 CPU 最大 8 个 DIMM 插槽，不能扩展内存

表 2-16 HP Integrity 入口级服务器提供领先的 2 路, 4 路指标

	HP Itanium 2 6M 1.6 GHz	AMD Opteron 1.8 GHZ, 64-bit	HP 快 1 倍
SPECfp_2000_base	#1 2119	1093	
SPECint_2000_base	#1 1322	1095	
SPECint_rate 2-way	#1 30.5	26.8	
SPECweb_SSL 2-way	#1 1930	1783	
SPECweb_SSL 4-way	#1 3702	3498	
SPECjbb 2000 2-way	#1 59,317	50,001	
SPECjbb 2000	#1 116,466	90,737	

4-way			
-------	--	--	--

表 2-17 HP Integrity 服务器与一些厂商 Opteron 服务器性能比较

指标类型	服务器类型	指标值	服务器类型	指标值
tpmC	rx2600	60,021	Newisys 2100	40 K
SPECfp2000_rate	rx2600	42.4	Newisys 2100	22.5
tpmC	rx5670	121,065	Newisys 4300	82,226

表 2-18 IPF 和 Opteron 平台上生态系统建设对比

	Itanium 2	Opteron
操作系统	HP 全面的多操作系统战略，在 IPF 平台上支持 HP-UX, OpenVMS, Linux 和 Windows 2003 等 64 位操作系统，所有主要的 Linux 厂商都已经提供 IPF 平台上的 64 位版本	只有 Suse 的 64 位 Linux 可以交付正式使用，其余的操作系统如 Microsoft, Red Hat 和 Solaris 等还在开发之中
应用软件	目前已有 400 多个 ISV 支持 IPF，已经有近 1000 个应用软件针对 IPF 的特点进行了优化。到 2004 年中，两者的数量将分别超过 1000 和 2000	只有很少的 ISV 支持 Opteron，达到同样数量的应用软件还需要花很大的力气；在 x86-64 上运行 32 位软件，操作码需要加上专门的前缀
子程序库	已经进行了支持近 1000 应用软件的共享子程序库的大量优化工作，数学核心 (MKL) 也已在 IPF 平台作了优化	水平相差很远
编译程序	Intel 和 HP 提供许多专门优化的编译程序，HP 今后单单通过使用新开发的编译程序就能提供 2 位数字性能增长	只有一般的第三方编译程序；在 Opteron 上运行 64 位应用，需要使用第三方编译程序，而它们的质量比 Intel 和 HP 专门开发的编译程序相差很远
驱动程序	拥有整套的 64 位操作系统下的驱动程序，驱动各种类型的设备	由于不能在 64 位操作系统下使用 32 位驱动程序，需要花很大努力才能驱动各种各样的设备

表 3-1 HP Integrity 系列中档服务器基本参数

	HP Integrity rx7620 8-路	HP Integrity rx8620 16-路	HP 服务器扩展部件
2-CPU 或 4-CPU 单元板	1-2	1-4	
1.3 或 1.5 GHz Intel Itanium2 处理器	2-8	2-16	
内存 (使用 512 MB, 1 GB 或 2 GB DIMM)	2-64 GB	2-128 GB	
热插拔 PCI-X I/O 插槽	15 插槽	16 插槽	16 插槽
聚合 I/O 插槽带宽	15.4 GB/s	15.9 GB/s	15.9 GB/s

PCI-X 插槽单总线带宽 (数量)	533 MB/s (1)	533 MB/s (2)	533 MB/s (2)
PCI-X 插槽双总线带宽 (数量)	1066 MB/s (14)	1066 MB/s (14)	1066 MB/s (14)
内部磁盘存储插槽/最大容量	4/584 GB	4/584 GB	4/584 GB
内部可移动介质插槽 (DVD, DAT)	1	2	2
硬件分区	2	4	
热交换冗余电源 (包括 N+1)	2	6	2
热交换冗余风扇 (包括 N+1)	Yes	Yes	Yes
高可用性	热交换冗余风扇和大容量电源 用于双电网保护的冗余电源线输入 对所有 CPU 和内存通道提供错误检查和校正 主内存 DRAM 自愈 (芯片备份) 奇偶校验保护 I/O 数据通路		
操作系统	HP-UX 11i v2; Windows Server 2003 Datacenter (只在 rx8620 上) Enterprise Edition (1Q2004; Linux Red Hat v3.0 (1H2004), OpenVMS (2005))		

表 3-3 HP Integrity rx7620 服务器主要竞争对手简表

	IBM	Sun	Bull	NEC	Dell
服务器	pSeries 650, 655	Sun FireV880, V1280	NovaScale 5080	Express5800 1080	PowerEdge 8450
处理器	Power4+ 1.45 GHz	UltraSparc III 1.2 GHz	Itanium2, 1.5GHz	Itanium2, 1.5GHz	Xeon MP 2,2.5, 2.8 GHz Xeon 3GHz
最大个数	8	8,12	8	8	8

表 3-4 HP Integrity rx8620 服务器主要竞争对手简表

	IBM	Sun	Bull	IBM	NEC	IBM
服务器	pSeries 670	Sun Fire 4800,680 0	NovaScale 5160	xSeries 455	Express5800 1160	xSeries 445
处理器	Power4+ 1.45 GHz	UltraSparc III 1.2 GHz	Itanium2, 1.5GHz	Itanium2, 1.5GHz	Itanium2, 1.5GHz	Xeon MP 2,2.5, 2.8 GHz Xeon 3GHz

最大个数	16	12, 24	16	16	16	16
------	----	--------	----	----	----	----

表 3-5 与 Integrity rx7620 相竞争的中档服务器系统资源

	HP Integrity rx7620	IBM pSeries p655	Sun Fire 4800	IBM eServer xSeries 445	Bull NovaScale 5080	Unisys ES7000 Aries 410
处理器	Itanium2 1.5GHz 或 1.3 GHz	Power4+1.5, 1.7 GHz	UltraSPARC III Cu, 1.2 GHz	Xeon MP 2,2.5, 2.8 GHz Xeon 3GHz	Itanium2 1.5GHz 或 1.3 GHz	Itanium2 1.5GHz 或 1.3 GHz
个 数	2-8	2,4,6,8	12	16	2-8	2-8
最大内存	2-64 GB	4-64 GB	2-96 GB	2-64 GB	4-64 GB	4-128 GB
最大 I/O 插槽数	15 个 64 位 PCI-X 插 槽： 其中 14 个 双带 宽, 1 个单 带宽	3 个 内 部 PCI-X 插 槽	8 PCI 插槽 (可扩展到 16 个插槽)	6 个 PCI-X 插槽	11 个 PCI-X	6 PCI-X
最大磁盘 容量	4 个港湾, 584 GB	36-2642 GB	无内部磁盘	292 GB	376 GB	-

表 3-6 与 Integrity rx8620 相竞争的中档服务器系统资源

	HP Integrity rx8620	IBM pSeries p670	Sun Fire 6800	IBM xSeires 455	Bull NovaScale 5160	Unisys ES7000 Aries 420
处 理 器	Itanium2 1.5GHz 或 1.3 GHz	Power4+ 1.5 GHz	UltraSparc III Cu 1.2 GHz	Itanium2 1.5GHz 或 1.3 GHz	Itanium2 1.5GHz 或 1.3 GHz	Itanium2 1.5GHz 或 1.3 GHz
个 数	2-16	16	24	16	16	2-16
最大内存	2-128 GB	4-256 GB	2-192 GB	1-224GB	4-128GB	4-256 GB
最大 I/O 插槽数	16 个 64 位 PCI-X 插 槽： 其中 14 个双带 宽, 2 个单 带宽	10 个内部 PCI 插槽	16 个 PCI 插槽	6 PCI-X	11PCI-X	6 个 PCI-X
最 大 内 部 磁盘容量	4 个港湾, 584 GB	36-7046 GB	无 内 部 磁 盘	1168 GB	576 GB	-

表 3-7 rx7620 是多种操作系统下性能最高的 8 处理器服务器

测试指标	处理器数	操作系统	指标	排名
SPECfp_2000_rate	4, 8	HP-UX	71.4, 142	8-路 #1
SPECint_2000_rate	4, 8	HP-UX	58.6, 116	8-路 #1
SPEC jbb 2000-java		HP-UX	190,349	2-路 #1
SAP SD 2 层	8	HP-UX/Oracle	用户数 1500	8-路 领先
SPECsfs97	8	HP-UX	71013	8-路 领先
SPEC web SSL	8	HP-UX	5,388	8-路 #1
SPEC OMPM 2001	4, 8		6555, 11098	8-路 领先

表 3-8 8620 是多种操作系统下性能最高的 16 处理器服务器

测试指标	CPU 数	操作系统	指标	排名
SPECfp2000_rate	8, 16	HP-UX	142, 234	超过 IBM 和 Sun 最佳的指标
SPECint2000_rate	8, 16	HP-UX	117, 232	16 路 #1
SPEC jbb 2000-java		HP-UX	341,098	4-路 #1
SPEC web SSL	16	HP-UX	9060	4-路 #1
SPEC OMPM 2001	8, 16	HP-UX	11847, 17852	16-路 领先

表 3-9 HP Integrity rx7620 8 处理器服务器提供领先的浮点和整数计算能力

厂商和服务 器系列	HP rx7620	SunFire	Fujitsu PRIMEPOWER	IBM xSeries	Bull Novascale	SGI Altix 3700
处理器	Itanium2	UltraSparc III	SPARC64	Xeon MP	Itanium2	Itanium2
SPECint2000	8 路 116	6800 16 路 122	850 8 路 79	445 16 路 131	5160 8 路 93	8 路 98
SPECfp2000	8 路 142	6800 16	850 8 路	-	5080 8 路	8 路

		路 153	110		125	142
--	--	-------	-----	--	-----	-----

表 3-10 HP Integrity rx8620 服务器提供领先的浮点和整数计算能力

厂商和服务 器系列	HP rx8620	SunFire	Fujitsu PRIMEPOWER	IBM pSeries	Bull NoveScale	Unisys ES7000 Aries
处理器	Itanium2	UltraSparc III	SPARC64	Power4+	Itanium2	Itanium2
SPECint2000	16 路 232	12K 32 路 232	1500 16 路 154	690 16 路 131	5160 16 路 117	420 16 路 137
SPECfp2000	16 路 234	12K 16 路 174	1500 16 路 194	670 16 路 187	5160 16 路 215	420 16 路 215

### 领先的 SPECweb99 SSL 联接指标

SPECweb99-SSL 基准测试测量服务器有效地处理安全加密的 Web 交易的能力，是企业用户在 Internet 上执行高度安全的交易所需的关键指标。HP 与领先的 Web 服务器基础设施供应商 Zeus Technology 公司合作，在 Integrity 平台上提供领先的安全加密 Web 服务解决方案。图 3-10 说明 8 路和 16 路 rx8620 都提供领先于 16 路 RISC 服务器的 SPECweb99-SSL 基准测试指标。

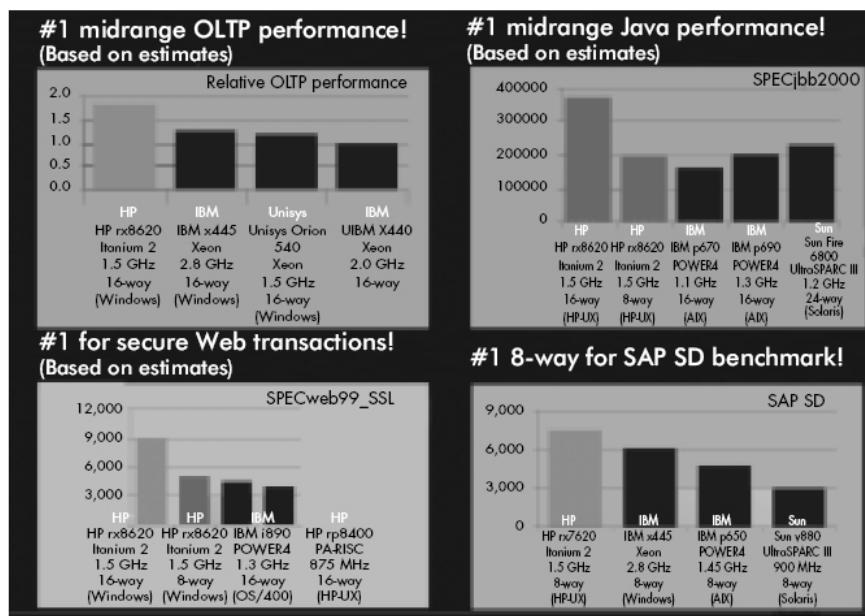


图 3-10 HP Integrity 中档服务器提供领先的性能

### 领先的 SAP 基准测试指标

SAP 是领先的 ERP 软件厂商, SAP R/3 是应用最广的企业应用软件。HP Integrity 中档服务器提供领先的 SAP 基准测试指标, 非常适合于满足中小企业数据中心管理和大型企业部门级应用。图 3-10 表示 rx7620 还提供领先的 SAP SD 基准测试指标。

### 领先的 SPECjbb2000 基准测试指标

SPECjbb2000 基准测试指标用来测试服务器一侧的 Java 性能, 提供测量服务器运行 J2EE (Java2 企业版)能力最客观和代表性的基准测试指标。运行 HP-UX Java 的 Integrity 中档服务器实现了业界最佳的 8 路和 16 路 SPECjbb2000 基准测试指标。图 3-10 说明 16 路的 Integrity rx8620 提供超过 16 路 IBM p670 和 p690 以及 24 路 Sun Fire 6800 的 SPECjbb2000 指标, 表明 Integrity 基于领先的支持 J2EE 框架下 Web Services 的性能。

### 全面领先和平衡发展的高性能

为了提供满足各方面需求的灵活性和全面的高性能, 现代的服务器产品不仅需要具有较高的单项指标(有些厂商的产品虽然某项指标较高, 但是有些指标却很差, 发展极不平衡, 这样的产品往往不符合支持企业数据中心应用的要求), 而且必须具有全面的高性能, 平衡的高性能指标, 才能更好地满足应用面较广的数据中心需求。图 3-11 和图 3-12 分别说明 HP Integrity 系列中档服务器具有领先于 IBM 和 Sun 对应产品的平衡性能。

#### 2.4.4 实现最佳的价格/性能

HP 在系统设计技术、支持多操作系统平台、连续可用性、高可伸缩性和可管理性等方面的优势, 大大降低了它基于 Itanium 产品系列管理、维护和升级费用、空间占用量、电源消耗和故障损失, 使之具有最佳的性价比和最低的总拥有成本。

### 最高的性能密度

计算机系统的性能密度描述达到规定性能占用的空间量。提高性能密度不仅能够缩小机房面积, 而且能够降低能耗、方便管理, 许多厂商都力图借此降低系统的总拥有成本。当前许多电信企业和网络服务器供应商采用大量中档服务器提供服务, 要求服务器占用尽可能少的空间。因此, 性能密度成为中档服务器的一个重要指标, 对于降低用户的总拥有成本具有重要的作用。HP 的优化设计技术使它的基于 Itanium2 的入口级服务器结构十分紧凑、提供远比竞争对手高的性能密度。由图 3-13 可见, 每个 HP 标准机架可以放置 4 个 rx7620 服务器、提供 672k tpmC 指标; 同样由图 3-14 可见, 每个 HP 标准机架可以放置 2 个 rx8620 服务器、提供 550k tpmC 指标, 而其他厂商同档次的服务器需要使用多得多的机架和服务器, 说明 HP Integrity 中档服务器提供最高的性能密度。

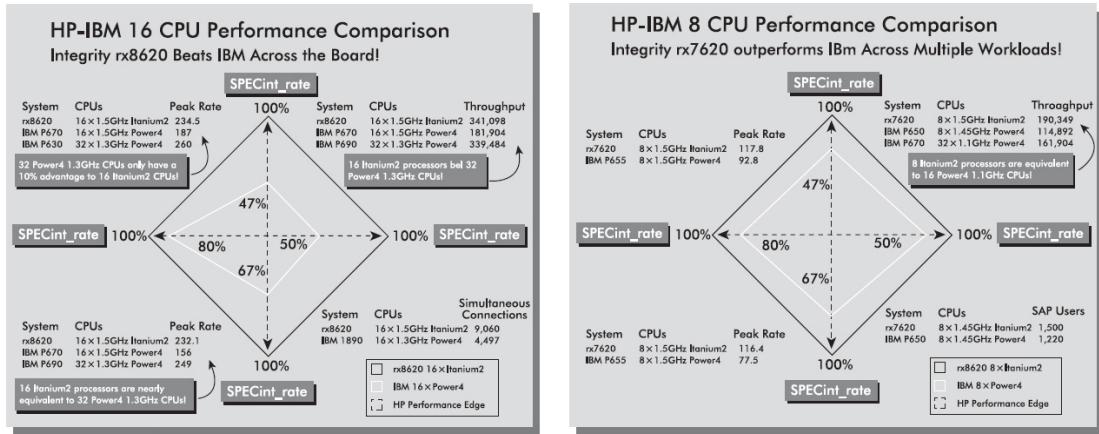


图 3-11 HP Integrity 系列中档服务器提供全面超过 IBM 对应服务器的性能

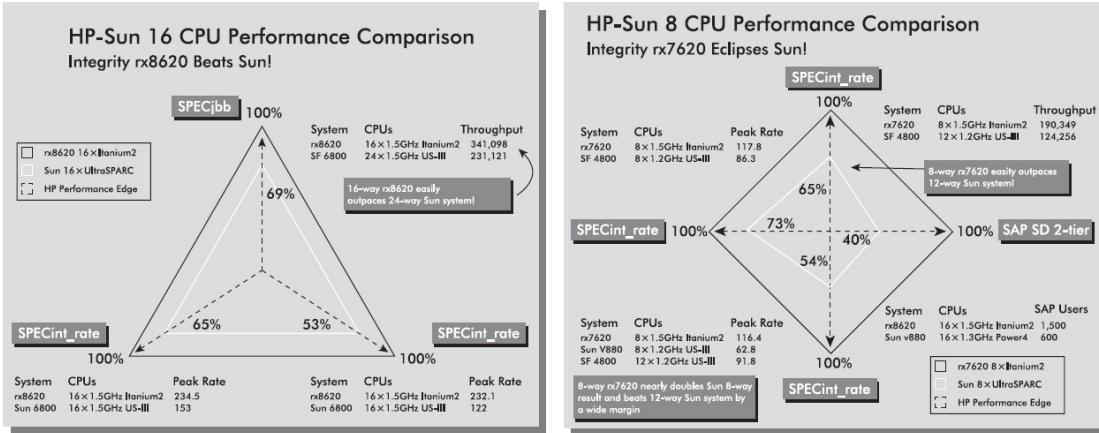


图 3-12 HP Integrity 系列中档服务器提供全面超过 Sun 对应服务器的性能



图 3-13 HP Integrity rx7620 提供领先的 8 路性能密度 图 3-14 HP Integrity rx8620 提供领先的性能密度  
最低的总拥有成本

随着企业间竞争的加剧，计算机系统的总拥有成本受到越来越多的重视。基于 Itanium 产品系列借助于 HP 的优势、提供最低的总拥有成本，使它们受到企业用户的广泛欢迎。

## 四、HP Integrity 系列企业级服务器

企业级服务器的一般特征是：装备 16 个以上高性能 CPU、价格不超过 300 万美元、在企业级层次应用。当前，随着 Internet 和电子商务爆炸性发展，企业级服务器在企业应用发挥着越来越大的关键作用，成为企业 IT 应用的核心。HP 基于 Itanium2 的 Integrity 系列目

前提供配置 16, 32 和 64 个 CPU 的 Superdome 企业级服务器。HP 通过全面贯彻其在工业标准部件基础上进行增值的战略, 使 HP 基于工业标准 Itanium2 的企业级服务器在满足企业用户需求方面具有全面领先的性能和特性, 成为企业用户最佳的选择。本节介绍 HP Integrity 系列企业级 Superdome 服务器的产品概貌、架构、特性和竞争优势。

## 4.1 HP Integrity 系列企业级服务器概述

HP Integrity 系列提供配置 16, 32 和 64 个 CPU 的 Superdome 企业级服务器, 其硬件的基本参数如表 4-1 所示。

## 4.2 HP Integrity 系列企业级服务器架构

HP Integrity Superdome 服务器的基本设计思想是: 采用适合于企业级资源规模和可伸缩性的缓存一致的非均匀内存访问 (ccNUMA) 架构和两层交叉交换互联拓扑, 采用 HP 领先的 sx1000 芯片组实施上述设计 (详见 1-3)。

### 4.2.1 互联拓扑

HP Integrity Superdome 服务器系统有三类基本组件: 单元或单元板, 交叉交换背板, 以及基于 PCI-X 的 I/O 子系统, 通过两层结构的交叉交换网络联接成一个完整的系统。图是这个系统的图示。

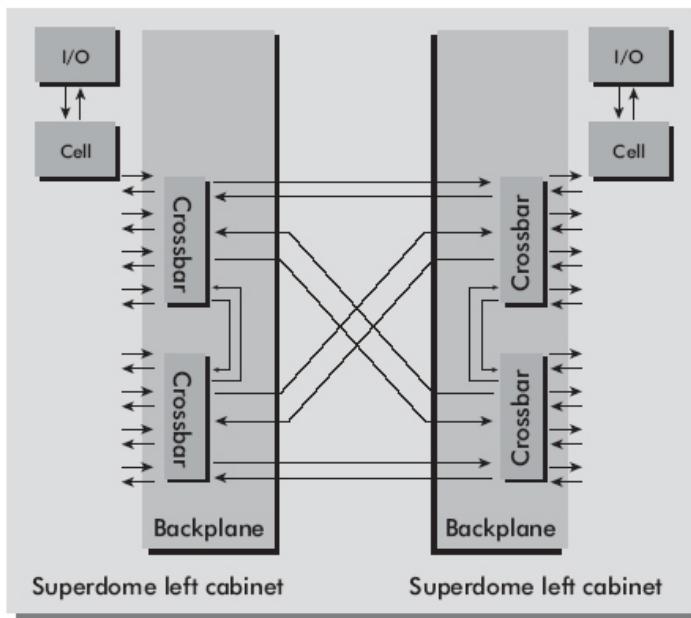


图 4-1 : HP Integrity Superdome 服务器的两层交叉交换互联拓扑

- 交叉交换网络组成方式
  - 总共 4 个交叉交换器
  - 每个交叉交换器联接 4 个单元
- 所有链路带宽和延迟时间相同, 实现:
  - 最小的延迟时间
  - 最大的可用带宽
- 网络完成点到点包过滤和路由, 确保信息安全和无阻塞传递, 同时实现隔离出故障的硬件和链路

- 每个服务器最多包含 16 个单元, CPU 与内存通信有 3 个延迟级别, 构成一个 NUMA 架构的系统
  - 本地内存: 与访问 CPU 在同一单元中的内存;
  - 本地交叉交换器内存: 与访问 CPU 在通过同一交叉交换器联接的单元中的内存;
  - 远程 crossbar 结构: 与访问 CPU 在通过不同交叉交换器联接的单元中的内存;

## 4.2.2 单元板

单元板是 Superdome 服务器的基本组成构件。每个单元板都是一个独立的部件, 包括: 对称多处理器 (SMP)、主内存以及所有必需的硬件:

- 最多 4 个处理器模块;
- 单元控制器 ASIC(专用集成电路);
- 主内存 DIMM, 最大容量为 32GB(每个板最多包含 32 个 DIMM, 4 个 DIMM 为一增量, 使用 512 MB 或 1 GB DIMM 或者两者的组合);
- 稳压器模块(VRM) ;
- 数据总线;
- 与 12 个 PCI-X I/O 插槽相连的可选链路;

Superdome 单元板内部也使用交叉交换器(称为单元控制器)把 CPU 模块、内存模块以及 I/O 适配器联接在一起。图 4-2 显示 Superdome 单元板的内部联接拓扑。每个单元拥有 16 GB/s 的最高内存带宽, 每个单元可选择是否与 12 插槽 PCI-X 卡护笼相接, 联接链路的最高带宽为 2 GB/s。每个单元到 crossbar 的带宽是 8 GB/s。

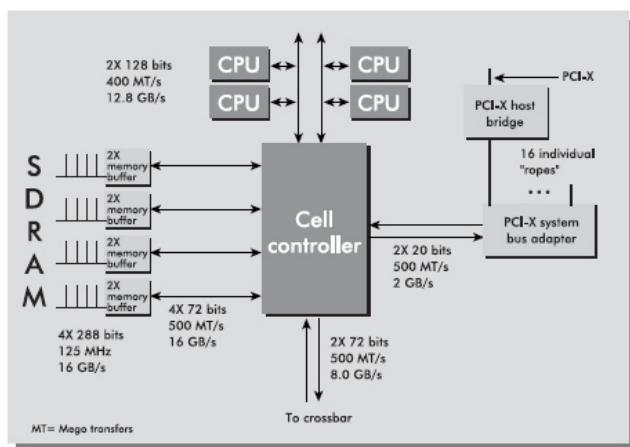


图 4-2 Superdome 服务器单元板内部互联拓扑

单元板上的单元控制器 ASIC(专用集成电路)是 Superdome 服务器系统 sx1000 芯片组的组成部分, 它负责协调单元板主要组件之间的通信, 确定一个请求是否需要与另外一个单元或 I/O 子系统通讯。

## 单元控制器 ASIC

单元控制器 ASIC 有 5 个主要接口:

- 4 个内存子系统;
- 2 个处理器端口(每 2 个处理器 1 个专用端口);
- 交叉交换器接口，通过这个接口与其它单元通信;
- 处理器依赖硬件 (PDH) ;
- I/O 接口，连接单元和 I/O 子系统

除了接口逻辑之外，单元控制器 ASIC 还保持整个系统上缓存的一致性。单元控制器 ASIC 同时支持 Intel Itanium 2 和下一代的 PA-RISC 处理器。处理器依赖硬件 (PDH) 是为单元复位提供所需本地资源，并将其带到可以加入其它单元的地点，引导操作系统的模块。PDH 包含系统引导固件，这个固件也是在运行时使用。

## **内存控制器 ASIC**

内存控制器 ASIC 也是 sx1000 系统芯片组的组成部分，它的主要功能是多路传输和多路分解单元控制器 ASIC 与内存子系统中 SDRAM 之间的数据。当单元控制器 ASIC 在内存接口命令总线上发布一个只读操作时，内存控制器 ASIC 缓冲 DRAM 读取数据，并尽快送回。当单元控制器 ASIC 发布写操作时，内存控制器 ASIC 接收来自单元控制器 ASIC 的写数据，并将其转给 DRAM。

注意，只有内存子系统的数据部分通过内存控制器 ASIC，所有的 DIMM 地址和控制信号都由单元控制器 ASIC 产生，然后通过内存接口地址总线直接发送给 DIMM，从而缩短了内存的延迟时间。

内存子系统具有四通路的通信机制，支持内存 DRAM 容错，即使一个独立的 SDRAM 芯片发生故障，也不会影响数据的完整性。内存子系统为单元控制器 ASIC 提供 16 GB/s 的峰值带宽，并将典型情况下与目录一致性相关的内务操作最小化，不仅如此，内存子系统从单元到本地内存访问的延迟时间也非常短：加载使用时的平均空闲延迟时间只有 245 ns。

### **4.2.3 交叉交换器背板**

每一个交叉交换器背板包含两组双交叉交换器 ASIC，通过这两组双交叉交换器 ASIC 在 8 个单元及其它背板之间提供无阻断的连接。每个背板机柜可以支持多达 8 个单元或 32 个处理器(在单机柜的 HP Integrity Superdome 服务器 32 路系统上)，两个背板可以用软电缆连接在一起，形成一个能够支持多达 16 个单元或 64 个处理器的机柜(双机柜的 HP Integrity Superdome 服务器 64 路系统)。

## **交叉交换器 ASIC**

交叉交换器 ASIC 是 sx1000 芯片组的另外一个组成部分，它实施高性能 8 路无阻断交叉交换通信机制和 500 MHz crossbar 链路协议，所有端口在功能和电规格上都完全相同。Superdome 服务器拥有一个充分连接、无阻塞交叉交换网络，共有 4 对交叉交换器 ASIC，每 4 个单元上有一个交叉交换 ASIC 对。

交叉交换网格的一个非常重要的特性，是所有的链路具有相同的带宽和延迟时间。这对

于最大限度地提高系统总体聚合带宽和最大限度地减少系统总体延迟时间具有十分重要的意义。单元到交叉交换器和交叉交换器到交叉交换器的通讯以相同的速度进行，从而控制访问远程内存的延迟时间，缩小访问本地和远程内存的延迟比。此外，Superdome 服务器的内存首先在单元之间、然后在内存条之间交错，这种交错设计可以平衡所有链路之间的内存通信量。

Superdome 的全程交叉交换器网络实现了一个全局的点到点包过滤网络，这个网状结构具有极高的完整性，每一个交叉交换通路完全独立。全程交叉交换器网络具有专用的数据和控制路径，每个通路可以完全独立于其它通路进行复位、分配或重新配置。Superdome 服务器的这一设计为资源隔离奠定了良好的基础。

交叉交换器 ASIC 提供一系列有助于提高 Superdome 服务器高性能的特性：

- 支持扩展到 128 路一致共享内存系统(采用 PA-8800、PA-8900 和 mx2 处理器)；
- 250 MHz 运行速度；
- 500 兆次传输/秒 (MT/s) 的链路速度；
- 链路协议支持两个交错式通道；
- 支持 Intel Itanium 处理器家族的双倍长度数据包模式；
- 性能计数器便于软件优化；

交叉交换器 ASIC 上的每个通路峰值带宽为 8 GB/s，这些通路为单元及其它交叉交换器 ASIC 提供高吞吐量路径：

- 交叉交换 ASIC 上的 4 个通路与 4 个单元连接(每个单元 1 个通路)。
  - 3 个通路连接到其余的 3 个交叉交换器 ASIC 上(在 64 路 Superdome 服务器系统中)。
- 每个 HP Integrity Superdome 服务器机型的总交叉交换带宽计算如下：  
(单元数量 × 每个单元的峰值交叉交换器带宽) ÷ 2 个通路  
各档 HP Integrity Superdome 服务器系统的 crossbar 带宽十分出色：
- HP Integrity Superdome 服务器 16 路系统的交叉交换带宽是 16 GB/s
  - HP Integrity Superdome 服务器 32-way 路系统的交叉交换带宽是 32 GB/s
  - HP Integrity Superdome 服务器 64 路系统的交叉交换带宽是 64 GB/s

## 内存和背板延迟时间

HP Integrity Superdome 服务器的设计可降低内存和背板的延迟时间，确保最优的性能。HP Integrity Superdome 服务器系统内的内存延迟时间有三种类型：

- 单元内的内存延迟时间是指运行在只包含一个单元板的 nPartition 上应用程序延迟；
- 同一个交叉交换结构上单元之间的内存延迟时间是指包含位于同一个交叉交换结构上的 4 个单元的 nPartition 延迟。例如，如果 nPartition 上有 4 个单元，则 1/4 的请求进入处理器所在单元板的内存，3/4 的请求进入其它 3 个单元板的内存；
- 不同 crossbar 结构上单元之间的内存延迟时间是指包含不全部在同一个 crossbar 结构上的由单元板构成的 nPartition 延迟。例如：如果 nPartition 上有 16 个单元，则 1/16 的请求进入处理器所在单元板的内存，3/16 的请求进入其它 3 个单元板的内存，最后，其余的 12/16 请求跨越两个 crossbar 结构；

HP Integrity Superdome 服务器的内存延迟时间取决于 CPU 的数量及相应单元板的位置。假定所有内存控制器都有平均分配的通信量，并且单元板的安装追求延迟时间最短，则平均的内存空闲延迟时间(系统上不执行任何应用程序)和内存延迟时间(加载到使用)显示如下：

#### 4.2.4 I/O 子系统

每个 Superdome 服务器单元都有可选的通向 I/O 机箱的链路：这样可以增强模块化，同时意味着不必在处理器的规模、内存和 I/O 之间权衡。每个单元通过一个 I/O 电缆链路与其远程 I/O 机箱相连。

HP Integrity Superdome 服务器的 I/O 子系统具有充分满足当前及未来扩展需要的能力。每个 I/O 模块包含 12 个 PCI-X 连接，包括 8 个标准 PCI-X 和 4 个高带宽 PCI-X 插槽，以及 1 个 I/O 控制器 ASIC 和电源。每个 PCI-X 插槽都有自己的 PCI-X 总线标准 PCI-X 插槽带宽为 533 MB/s，高带宽 PCI-X 插槽则达到 1066 MB/s 的带宽。点到点的连接可以尽早检测、抑制和校正错误。

任何 I/O 模块都可以支持一个核心 I/O 卡(是每个独立的 nPartition 所需的)。HP Integrity Superdome 服务器 16 路、32 路和 64 路系统在系统机柜内可以分别容纳 4、4 和 8 个 I/O 模块，总 PCI 插槽数分别达到 48、48 和 96 个。也可以添加一个 I/O 扩展柜。在 HP Integrity Superdome 服务器 32 路系统上，I/O 扩展柜可增加 48 个 PCI-X 插槽，使 8 个单元的最大连接能力达到 96 个 PCI-X 插槽。对于 HP Integrity Superdome 服务器 64 路系统，I/O 扩展柜则可增加 96 个 PCI-X 插槽，使 16 个单元的最大连接能力达到 192 个 PCI-X 插槽。

配置 16 个单元的系统 - 每个单元有自己的 I/O 模块和核心 I/O 卡 - 可以支持高达 16 个独立 nPartitions。注意：单元可以在不连接 I/O 模块的情况下配置，除非连接单元，否则在系统上不能配置 I/O 模块。

I/O 子系统带宽是每个单元 2.0 GB/s，这样，HP Integrity Superdome 服务器 16 路、32 路和 64 路系统的 I/O 子系统总带宽分别可达到 8.0 GB/s、16.0 GB/s 和 32 GB/s。

### 4.3 HP Integrity 系列企业级 Superdome 服务器特性

HP Integrity 系列 Superdome 服务器采用与提供一系列支持现代企业级服务器所具备的特性，满足企业支持关键任务应用、高性能技术计算等方面的需求。

#### 4.3.1 高可用性

Superdome 提供高于入门级和中档服务器的单系统高可用性，包括系统可靠性、可支持性和可维护性，以及提供建立集群提供更高的可用性。

##### 系统可靠性

Superdome 服务器采用高度可靠的部件和工艺、先进的设计水平和生产技术大大提高了系统的可靠性，包括：

- 高度可靠的部件和制造工艺：ASIC 完全烧入和高质量生产工艺，使 ASIC 故障率改进了 10 倍；所有的 HP 动能 超腾服务器关键组件上进行完全的故障测试和寿命加速测试大大提高其可靠性；
- 全面的故障探测和校正：对整个系统提供全面的奇偶校验保护，包括 DIMM 地址、CPU 总线、交叉交换器和 I/O 链路；提供强大的 ECC 保护功能，能够自动校正内存、CPU 缓存等关键部件的 1 位错误、能够探测多位错误；
- 强大的动态自愈功能：当内存 DIMM 过于频繁出现故障时，自动将坏内存页面去分配，

不再参与运行；当发现 CPU 出故障后，系统将自动停止把新的进程分配给该 CPU。从而保持系统的正常运行；当系统重新启动时，已经标记为故障和去分配的内存、CPU 和交叉交换器链路等部件不会被重新投入使用、以免影响系统的正常运行；

- 容错和错误恢复功能：支持内存 DRAM 容错、I/O 错误恢复和 I/O 卡故障的系统恢复、I/O 到单元控制器链路的故障恢复、交叉交换器线路故障的恢复；
- 完全的硬件故障隔离功能：提供 nPartition、I/O 卡相互完全隔离，当一个硬件分区或 I/O 卡发生故障不会影响其它分区或 I/O 卡的正常工作；

表 3-11 HP Integrity 服务器具有最低的总拥有成本

	IBM pSeries 670	HP Integrity rx8620	节省	
采购价格	\$318,536(8-路, 1.5 GHz)	\$209,760(8-路, 1.5 GHz)	\$104,775	33%
每次操作价格	\$2.08/操作	\$0.75/操作	\$1.33/操作	64%
支持费用	\$80,312	\$71,895	\$8,617	11%
升级费用	\$72,482 (硬件升级) Power5 不能机箱内升级	\$25,100 (硬件升级) HP mx2 CPU 可以在机箱内升级	\$47,382	65%

表 4-1 HP Integrity 系列企业级 Superdome 服务器基本参数

	16 路 Superdome 服务器	32 路 Superdome 服务器	64 路 Superdome 服务器
2-CPU 或 4-CPU 单元板	1-4	1-8	3-16
Intel Itanium 2 1.5 GHz 处理器	2-16	2-32	6-64
内存 (使用 512 MB 或 1 GB DIMM)	2-128 GB	2-256 GB	6-512 GB
12 插槽 I/O 卡护架	1-4	1-4 IOX 为 1-8	1-8 IOX 为 1-16
热插拔 PCI-X I/O 插槽	48 个插槽 (32 个标准 PCI-X 插槽, 16 个高 BW PCI-X 插槽)	48 个插槽 IOX 为 12-96 个(64 个标准 PCI-X 插槽, 32 个高 BW PCI-X 插槽)	96 个插槽 IOX 为 12-192 个(128 个标准 PCI-X 插槽, 64 个高 BW PCI-X 插槽)
热插拔冗余电源(含 N+1)	4	6	12
I/O 风扇	6	6	12
热插拔冗余风扇 (含 N+1)	4	4	8
I/O 带宽 (峰值)	8 GB/s	16 GB/s	32 GB/s
单元控制器到内存子系统带宽 (峰值)	16 GB/s		

I/O 带宽 (峰值)	8 GB/s	16 GB/s	32 GB/s
标准 PCI-X 总线带宽	533 MB/s		
高 PCI-X 总线带宽	1066 MB/s		
2X PCI I/O 总线带宽	266 MB/s		
4X PCI I/O 总线带宽	533 MB/s		
单元控制器到 I/O 子系统带宽 (峰值)	2.0 GB/s		

表 4-2 Superdome 访问内存延迟时间		
单元板数量	CPU 数量	平均的空闲加载到使用内存延迟时间
1 个	4 个	246 ns
2 个	8 个	330 ns
4 个	16 个	371 ns
8 个	2 个	3 417 ns
16 个	64 个	440 ns

## 可支持性

可支持性是 HP Integrity Superdome 服务器的另外一个重要特性。Superdome 服务器的部分可支持特性包括：

- **事件监测服务 (EMS) (仅在 HP-UX 上可用):** HP-UX 上的 HP 事件监测服务(HP Event Monitoring Service, EMS)软件通过跟踪系统的重要信号，辅助系统管理，监测系统上的几乎所有硬件，包括：大容量存储器、内存、光纤通道组件(多路复用器，交换机，卡，光纤等等)、I/O 卡、主系统总线上的 ECC 错误、CPU 缓存中的 ECC 错误、系统温度、处理器本身、有关硬件和电池状态、机柜风扇、机柜电源、机箱代码日志故障等。此外，它另外还能够监测系统硬件配置和选择的内核参数。当硬件检测到问题时，就产生一个事件，并通过 SNMP、OpenView Vantagepoint Operations、电子邮件、页面、系统日志、控制台或者选定的文本日志文件向管理人员报告。每个事件包含对问题的全面描述，严重性分类(信息，警告，严重，危急)，以及显示可能的原因和建议措施的文本。如果检测到过多的处理器缓存错误，处理器即自动停止使用，直到更换。如果客户在分区的 HP-UX 上有可用的 iCOD 处理器，那么 EMS 就会自动启动其中的一个处理器，替换已停止使用的处理器。另外还将故障处理器做上标记，系统下一次重新启动时将不再使用它；
- **改进的支持工具管理器软件(STM):** 为管理人员提供有关系统的详细信息。HP 动能 超腾服务器的设计使每一个现场可更换单元 (FRU) 都能够报告诸如序列号、部件号、修订

级等信息。STM 的系统信息工具使管理人员能够方便地看到这些信息，以便进行管理和维修。此外，管理人员也可通过桌面管理接口 (DMI) 访问硬件清单信息；

- **支持管理站 (SMS)**：为了进一步降低意外停机的可能性，Superdome 服务器提供一个独立的服务器。支持管理站为 Superdome 系统提供诊断和测试功能，包括对专门设计的集成电路状态进行扫描的工具。HP 支持工程师可以使用 SMS 在数据中心对 Superdome 服务器系统进行诊断；
- **即时支持企业服务 (ISEE)**：ISEE 是一个领先的远程支持解决方案，它与相应的程序和支持人员相结合，为关键任务运行环境强大的即时支持能力。HP ISEE 解决方案通过经常性、自动地收集客户的关键任务环境数据、远程诊断和主动服务等方式大大提高 Superdome 服务器系统的可用性；
- 

### 可维修性

HP Integrity Superdome 服务器具有大量使其维修更方便的特性，维修时只需很少时间甚至无需停机。这些特性包括：

- **可按需即时增容的 (iCOD) N+1 CPU 和单元板 (只在 HP-UX 上可用)**：Superdome 服务器在 HP-UX 下支持按需即时增容 (iCOD)，这个特性可以联机添加 CPU 或单元板，无需系统重新启动。因此，增加容量不会影响系统的可用性。iCOD 的另外一个优点是允许建立 N+1 个 CPU 或单元板的系统，确保最大的单系统可用性。如果一个 CPU 或一个单元板(例如 CPU 或内存)发生故障，随时可以自动接替，不必中断系统的正常运行；
- **热插拔 N+1 风扇、电源和背板 DC/DC 转换器**：Superdome 服务器配有 N+1 风扇、电源和背板 DC 转换器，从而确保这些组件的最大可用性；
- **联机更换 PCI-X I/O 卡**：运行 HP-UX 的 Superdome 服务器支持允许联机添加和更换 (OLAR)PCI-X I/O 卡，运行 HP-UX 和 Windows Server 2003 的 Superdome 服务器支持 PCI 卡 OLAR 特性，使管理人员可以在系统运行过程中、不影响其它组件或不重新启动的情况下，添加新卡和更换旧卡；
- **联机添加/更换单元板 (只在 HP-UX 上可用)**：Superdome 服务器支持联机添加和更换单元板，可以在系统不停机的情况下对这些关键组件进行维修和维护；
- **联机添加 nPartitions**：管理人员在添加 nPartitions 时不影响其它正在运行的 nPartitions，这样的动态重新配置是 Superdome 服务器能够提供超长运行时间的主要原因之一；
- **双电源**：Superdome 服务器上的双电源意味着电源可以得到保护，不会成为单点故障；

为了进一步增加应用程序在 Superdome 服务器系统内的运行时间、提高可用性，可以把 Superdome 的多个硬件分区 nPartitions 组成 Serviceguard HA 集群系统。检测到 nPartition 内的故障以后，Serviceguard HA 集群软件将发生故障的分区中的应用程序转移到 Superdome 服务器系统内的另外一个 nPartition 上，继续运行。

### 4.3.2 服务器虚拟化

HP 在 Superdome 服务器上提供基于领先的分区技术的服务器虚拟化功能，使得 Superdome 具有更强的支持企业应用特性，全面满足企业用户对企业级服务器的各种需求，促进了 Superdome 在企业中的广泛应用。

服务器虚拟化使管理员可以将单或多服务器环境配置为一个可重用的资源池，从而优化

使用，简化管理。虚拟化意味着物理资源与服务器基础设施架构的逻辑视图分离。分区是隔离一个或多个服务器内操作环境的物理或逻辑机制，分区为 IT 经理提供动态调整应用程序资源使用量的灵活性，同时保证所有的应用程序都能够免受可能导致服务中断或性能降级的中断事件的影响。HP 的连续分区解决方案提供多种硬、虚拟和资源分区工具，在服务器或分区级实现资源的虚拟化，提高系统和子系统的总体利用率，降低整合环境下的成本。

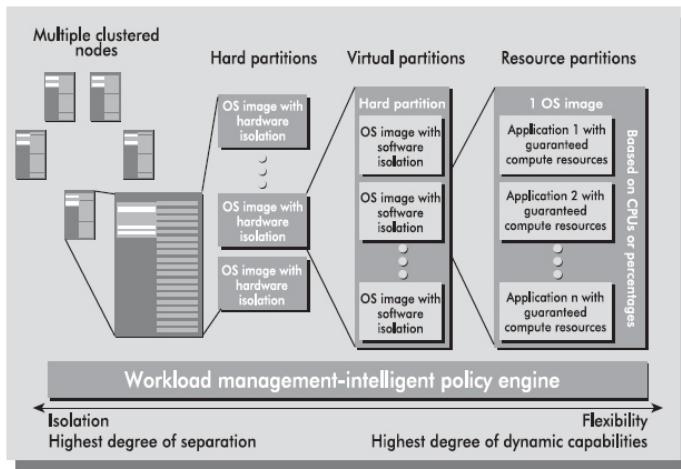


图 4-3 HP Partitioning Continuum 将高度隔离与出色的灵活性结合在一起

HP 提供四种不同类型的分区解决方案，每一种类型支持不同的应用程序隔离和资源优化平衡。

## 硬件分区

结点内的硬件分区用于将应用环境与单路径故障点 (SPOF) 隔离，这意味着在硬件分区 内运行的应用程序不受其它分区所发生的硬件或软件事件的影响。在 Superdome 服务器上，硬件分区在结点内，称做 nPartition 或 nPar。Superdome 服务器可以支持从 1 个到 16 个之间任意数量的 nPartition，每一个都将支持自己的操作系统、应用程序、外设和网络。nPartition 可以安装 HP-UX、Linux 和 Windows 操作环境。

在 HP 动能超腾服务器内，单元板划分成物理结构，一个 nPartition 包括一个或多个单元，这些单元在高带宽、低延迟 crossbar 结构上协调地通讯。单元板内特殊的可编程硬件定义 nPartition 的边界，强制与其它 nPartition 的动作相隔离。每一个 nPartition 都运行自己独立的操作系统，不同的 nPartition 可以执行相同或不同版本的操作系统，甚至不同的操作系统(例如 HP-UX、Linux 和 Windows)。

每一个 nPartition 都有自己独立的 CPU、内存和 I/O 资源集，管理人员可以使用系统管理命令，在不从物理结构上改变硬件的情况下，将资源从一个移到另外一个 nPartition 上。此外还支持动态添加新的 nPartition。

## 虚拟分区(HP-UX 11i v3)

虚拟分区提供一个服务器或硬分区内的全面软件故障隔离，这意味着任何与应用程序或

操作系统相关的故障都只影响它正在执行的分区 - 对同一个系统上运行的其它虚拟分区没有任何影响。在采用虚拟分区的系统上，每一个操作系统实例都完全独立于其它的所有操作系统。不同分区上的操作系统可以是不同的版本，也可以有不同的调整参数，因此，虚拟分区对于测试新的操作系统版本或应用程序十分有用。

## 资源分区

资源分区适应在一个操作系统实例内相互竞争的应用程序之间动态分配专用资源的需要，以避免资源争用。HP 为 HP-UX 环境提供 HP Process Resource Manager，它允许系统管理员控制应用程序、用户或组在高峰系统负荷时可使用的资源量。对于 Windows 环境，Microsoft 的 Windows System Resource Manager (WSRM) 能够提供资源分区。

HP 提供多个虚拟化解决方案、支持在 Superdome 服务器上建立高可用的垂直扩展环境。在垂直扩展的服务器上，分区可以与 HP 按需解决方案相互补充，为实施计算能力的动态扩展和降低所需的基础设施架构硬件及软件提供众多选项。HP HP-UX 服务器的具体解决方案是按需即时增容 (iCOD) 和按使用付费 (PPU)。iCOD 允许客户在需要时启动一个分区或服务器内的处理器(处理器也可以临时启动 [ TiCOD]，以满足短期需要，然后停用并再次保留)。HP PPU 解决方案是基于使用的租用解决方案——客户只为实际使用的资源付费。就虚拟化而言，HP 按需解决方案将费用直接与特定 IT 服务实际使用的资源相匹配，只在真正需要时启动。垂直扩展一般意味着使用几个整合的高性能服务器同时运行很多复杂的应用，这个环境为降低成本和现有资源的更好利用提供了很多机会。HP 服务器分区和资源优化解决方案可以帮助管理员在保证服务级别不降级的情况下，将服务器利用率从典型的 15-50% 提高到 90% 以上。虚拟化服务器环境与按需即时增容 (iCOD) 和按使用付费 (PPU) 等按需解决方案的结合，使客户可以只在需要时启动附加的容量，根据实际使用情况购买服务器资源。

### 4.3.3 高可伸缩性

可伸缩性目前已经成为用户对企业级服务器系统的最基本需求之一。这要求服务器系统(包括它的硬件和软件资源)能够在保持硬软件兼容性的同时，通过向上扩展(即增加资源)提供更高的性能和更强的功能并且能够通过向下缩小(即减少资源)降低成本。

Superdome 服务器能够最经济、快速和有效地全面各种用户对可伸缩性的要求，包括提供最大的系统增长空间、满足爆炸性增长的应用需求；提供尽可能低入口点，允许用户只投资购买当前需要的设备、以节约初始投资，同时保留需要时再投资扩展的余地；在扩展过程中保持硬软件兼容性，保护用户原有的投资。

## 最大的增长空间

Superdome 服务器不仅当前能够提供最大系统资源，而且为进一步扩大系统资源留有最大的增长空间，允许通过采用新技术(如新一代处理器)、扩大机器的规模(如处理器个数)、添置更多的存储设备(高速缓存、主存和磁盘等)、改进软件等各种途径提高系统的性能或增加系统的功能，满足用户不断增长的需求：

- **提供更高的数据处理能力和内存容量**： Superdome 服务器通过 mx2 模块可以把处理器数量增加一倍(达到 128 个)、或者采用 IPF 以后各代性能更高的处理器提供更高的数

据处理能力；Superdome 通过采用密度更高的 4GB DIMM 成为业界第一个支持 2 TB 内存的服务器；

- **通过集群和系统互联扩大规模：**通过把 Superdome 服务器联接在一起组成 Hyperplex 集群系统、能够大大扩展系统的规模。事实上，TOP500 中有 126 套以上 HP9000 Superdome 服务器的集群系统。人们预期，基于 Integrity Superdome 的集群系统也将在 HPTC 应用的最高端发挥很大的作用；
- **可扩展的外设联接和网络通信能力：**Superdome 的支持 192 个 PCI-X I/O 端口和 32 GB/s IO 带宽能力与 HP 在 SAN、StorageWorks 和 GigaBit 交换器等方面的技术优势相结合，将使 Superdome 服务器在支持外设和网络通信方面具有无可估量的扩展潜力，满足企业信息中心发展的需要；
- **软件可伸缩性：**Superdome 服务器的空前容量以及康柏最成熟的 64 位技术(包括操作系统、中间件、VLM 等)为系统软件的发展、应用软件性能的提高提供最好的舞台。人们必将看到在基于 IPF 的 Superdome 平台上创造出比基于 PA-RISC 处理器系列的 HP9000 Superdome 平台上更加辉煌的成果；

## 最高的扩展效率

Superdome 服务器的可伸缩性优势还在于提供最佳的性能扩展线性、最低扩展成本和最快的扩展速度，从而实现最高扩展效率：

- **最佳的性能扩展线性：**Superdome 服务器不仅提供高系统互联带宽和内存带宽，而且能够实现随着系统的扩展始终保证为每 4 个处理器提供 12.8 GB/s 带宽、随着系统负载的增加访问远程内存与访问本地内存时间之比不超过 2 倍。因此，Integrity 系列服务器能够实现系统总的处理能力与资源的增加成正比，提供最佳的性能扩展线性，使得用户确确实实能够得到与投资增加成比例的性能扩展，满足工作的需要。表 4-3 中的数据说明 Superdome 服务器处理器的个数由 8 个增加到 16 个、32 个和 64 个是其浮点、整数、TPC-C 指标始终基本上保持与处理器个数同步线性增加；
- **最低的扩展成本：**Superdome 服务器向上扩展的成本是最低的。这不仅是由于系统模块化和标准化的设计而且也得益于 HP 强大的全球服务力量，使得用户能够以最合理的代价实现系统扩展；
- **最快的扩展速度：**Superdome 服务器在扩展时用户可以通过购买相应的扩展部件方便地现场升级带电升级、外设也可以热插拔，而不干扰系统的运行；HP 提供 i COD (资源立即按需供应) 选购件：HP 将在工厂中为购买这一选购件的用户把附加的资源如 CPU 和内存等预装在系统中。当用户需要使用它们时，在支付相应的费用后，HP 可以立即释放这些资源供用户使用，既不干扰正常操作也没有任何延迟！

## 最强的投资保护

Superdome 服务器在升级和扩展过程中，将尽可能保证相同的硬件、系统软件、应用软件仍然可以继续使用、几乎不必作任何修改，从而为用户提供最强的投资保护，包括：

- **在扩展过程中保护用户投资：**Superdome 服务器在扩展过程中，除了增加必要的新设备外，用户原有设备都可以使用，实现 100% 的扩展投资保护；

- 提供最大的机箱内升级空间：Superdome 服务器不仅能够实现机箱内现场向上升级，而且允许在系统混合使用装备不同主频处理器的单元，从而使得用户能够在继续使用原有单元条件下，在系统中加入装备更新、更快处理器的单元；以前的服务器等升级到 Superdome，也允许用户继续使用大量原有部件如磁盘、磁带机、互联设备和外围设备等；

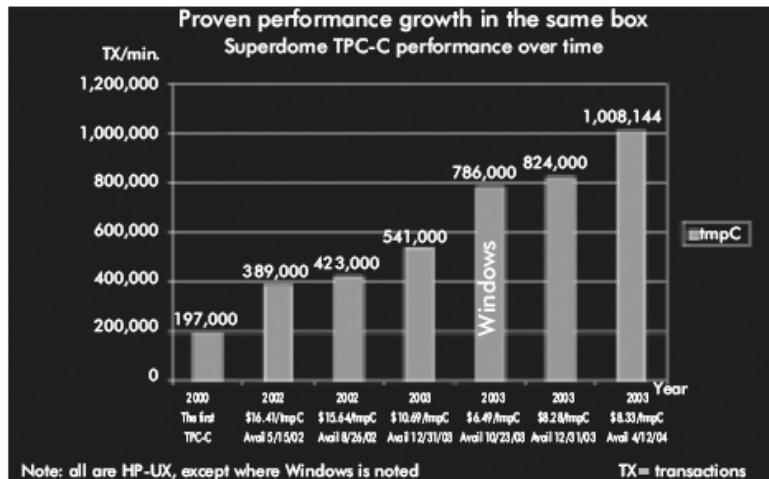


图 4-4 Superdome 提供最强的机箱内可扩展能力，把 OLTP 能力提高 5 倍以上

- 提供软件的二进制兼容性：保证应用软件在整个 Integrity 系列上二进制兼容是该系列的设计思想，用户现有的应用软件都可以不加任何修改地在 Superdome 服务器上以更高的性能运行，从而保护用户的软件投资；

## 4.4 HP Integrity 系列企业级服务器竞争优势

计算机问世的初期主要应用于科学计算，形成了完整的高性能技术计算应用领域。20世纪 70 年代以后，计算机应用逐步转向以信息处理为主，蕴育了规模更庞大的企业应用领域。当前，随着 Internet 的爆炸性发展和 eBusiness 技术的快速推广，这两类应用在技术上互相渗透、系统的规模越来越大、结构越来越复杂，对支持这些应用高端的企业级服务器也提出了越来越高级和全面的需求。如何挑选最合适的企业级服务器成为广大用户关心的焦点、而如何夺取更多的用户也成为各厂商竞争的热点。特别是由于企业级服务器在总投资中占有较大比例且在企业信息应用部门的核心地位，企业挑选其他各级服务器、存储设备、外设到网络通信整个 IT 基础设施的设备选型以至应用软件、技术服务和合作伙伴都往往受到企业级服务器选型的巨大影响。因此，开发企业级服务器新技术和创造重要测试指标新记录(如 TPC-C)成为各厂商努力攀登的高峰，企业级服务器的市场竞争成为厂商拼搏得最剧烈的角逐场地。

### 4.4.1 市场竞争态势

HP Integrity 系列 Superdome 服务器主要面对的是各厂商企业级 RISC 服务器的竞争，一些厂商生产的基于 Itanium2 的高端服务器也成为 16 和 32CPU 的 Superdome 服务器的竞争对手。

此外，还有 SGI Altix 3700 虽然也能够支持 64 个 Itanium2 处理器，但由于它的应用领域主要集中在高性能技术计算，主要与基于入门级 Integrity 服务器的集群系统竞争，而不构成 Superdome 的竞争对手。

## 2.4.2 工业标准平台上的技术优势

当前，用户对企业级服务器在系统设计技术上提出了更高的要求，其中最关键的是：

- 提供丰富的系统资源容量、支持更大规模应用和更广泛类型网上服务；
- 建立基于领先体系结构和互联拓扑的系统基础设施、以充分发挥庞大系统资源的潜力；
- 具有全面和经济可伸缩性，在保护原有投资的条件下提供最大的扩展空间、以满足工作发展的需要；
- 支持方便灵活的编程模式和丰富的应用软件；
- 提供企业信息中心所需的高可靠性、可用性和可维护性(RAS)；

HP 在工业标准技术基础上进行创新和增值的产品总体战略，奠定了 HP Integrity 系列的技术优势。对企业级服务器，这一战略体现在基于工业标准平台提供一系列能够满足上述技术要求、充分发挥企业级服务器资源优势的特性，从而在领先的技术优势基础上确保 HP Integrity 企业级 Superdome 服务器提供全面领先的性能、最佳的价格/性能，成为企业用户的最佳选择。

表 4-4 HP Integrity Superdome 服务器主要竞争对手简表

	IBM	Sun	Fujitsu	NEC	Unisys
服务器	pSeries 690 Turbo	Sun Fire 15000	PrimePower Model 2500	Express/1000 Model 1320 Xc	ES7000 Orion 430
处理器 类型和 主频	Power4+ 1.7/1.5 GHz	UltraSparc III Cu 1.2/1.05 GHz	SPARC64 V 1.35 GHz	Itanium2, 1.5GHz	Itanium2, 1.5GHz
处理器 个数	8-32	16-106	16-128	32	2×16

表 4-5 各厂商企业级服务器系统资源容量

厂商名称	HP	IBM	Sun	Fujitsu	NEC	Unisys
服务器	Integrity Superdome	pSeries 690 Turbo	Sun Fire 15000	PrimePower Model 2500	Express/1000 Model 1320 Xc	ES7000 Orion 430
处理器类 型和主频	Itanium2, 1.5GHz	Power4+ 1.7/1.5 GHz	UltraSparc III Cu 1.2/1.05 GHz	SPARC64 V 1.35 GHz	Itanium2, 1.5GHz	Itanium2, 1.5GHz
最大处理 器数	64	8-32	16-106	16-128	32	2×16
最大内存 容量	512GB	512GB	16-576 GB	2-512 GB	256 GB	128 GB

最大 I/O 插槽数	192 PCI-X	20 PCI	72 PCI	320 PCI	56 PCI-X 或 64 PCI	32 PCI-X
------------	-----------	--------	--------	---------	----------------------	----------

## 二、HP 基于 IPF 集群系统硬件

HP 基于 IPF 集群系统硬件主要包括集群节点(计算和管理节点)和互联系统(互联网络和通信协议)以及存储设备等。

### 2.1 系统互联架构

HP 集群系统的基本设计思路是通过高可伸缩、低延迟互联网络设备把大量装备 IPF 处理器的商品化服务器或工作站联接起来，在相应的集群系统软件和其他软件工具支持下构成一个提供并行编程环境的高端或超级计算机系统。由于超级计算机系统是由许多个独立节点通过网络联接而成的，内部互联网络设计已经成为确保系统高性能、高性价比和高可伸缩性的关键。

#### 集群内部互联网络设计要求

为了实现在商品化的服务器或工作站基础上构件超级计算机，关键是采用适当的互联架构和相应的系统软件。其中，互联网络架构是设计集群模式超级计算机系统的基础。集群系统的互联网络架构必须是无共享线路的一体化消息传递网络，同时满足高带宽、低延迟、无阻塞、高可伸缩性和经济性等要求。

#### 提供高带宽

互联网络带宽定义是，消息(一组具有指定格式的信息)进入网络后网络传输信息的最大速率。过去一般以 Mbps 或 Gbps 为单位。网络带宽又可以细分为：端口带宽和总带宽(聚合带宽)。端口带宽指网络任何端口每秒能够传送到任何其他端口的最大位数；总带宽定义为每秒能够从一半节点传送到另一半节点最大位数。为了支持超级计算机系统，系统互联网络必须具有高带宽，包括总带宽和端口带宽。Gbps 已经成为端口带宽起码的基线指标。聚合带宽目前正在向 Tbps 方向发展；

#### 实现低延迟

网络通信延迟指从源节点向目标节点发送一个消息所需的时间总量。这一延迟由四个时间分量组成：

- 软件开销：网络两端发送和接收消息所花费的时间，包括发送节点把消息放到网络上的所花费的时间和接收节点把消息从网络上取出来的所花费的时间；
- 通道延迟：通道被占用所花费的时间(等于消息长度除以通道带宽)。通道延迟通常由瓶颈联线或通道确定；

- 路由延迟：消息传输过程中在通过的各个交换器时进行路由决策所花费的时间。路由延迟与路径距离(路径长度或两端间中继站数)成正比；
- 竞争延迟：网络中的消息传输冲突所花费的时间，它依赖于网络上的交通条件；  
集群系统内部互联网络必须具有尽可能低的通信延迟，方能确保提供所需的高性能。

## 实现无阻塞消息传递

网络通信的方式可以设计成有阻塞或无阻塞的。有阻塞网络允许在通道由于传递多个消息而发生冲突时把消息暂时存储在缓冲区中以后再发送。因此，有阻塞网络也可以称为有缓存的网络。在无阻塞网络中，不使用缓冲区，而使用允许网络选择另一个路径的方式来解决消息间的冲突。集群系统内部必须使用无阻塞网络在节点之间传递消息。

## 高可伸缩性

由于超级计算机系统是由许多个节点通过内部互联网络联接而成的，因此要求其内部互联网络能够从支持较少的节点出发、扩展到支持尽可能多的节点，并且在扩展过程中保持原有的总体结构稳定不变、不降低原有的端口带宽、提供更高的聚合带宽，从而确保满足超级计算机系统可伸缩性的需要。

## 经济性

当前由于超级计算机系统已经不再是少数机构垄断的贵族化产品，其内部互联网络的设计也必须体现经济性的原则，注意高性能和低成本相结合，确保最终产品的性价比。

## 交叉交换互联架构

系统内互联网络通常有三种架构：平布总线、完全互联交叉交换器和树结构。这三者各有其优点和缺点(详见表)。

为了提供支持超级计算机系统所需的高带宽、低延迟，总线(如标准的以太网或高速以太网)和树结构显然都不能满足设计要求，超级计算机系统的内部互联网络必须采用交叉-架构。图 4 表示所使用的 8 端口交叉交换器，满足了高带宽、低延迟和无阻塞的要求。但是，它的可伸缩性很差，当端口数增加时其成本将急剧增加，经济上也难以承受。

在目前技术水平下，使用 8 端口交叉交换器既能够联接相当大的资源量、实施起来比较简单、代价也不太高。16 个端口已经几乎是技术和经济上可接受的上限。

## 混合的互联架构

为了支持集群系统必须在交叉交换器基础上进一步改进。较好的策略是采用混合架构，即把它们结合起来混合使用。表说明常用的系统互联混合架构。

HP 基于 IPF 的集群系统采用交叉交换+ 树结构的胖树模式进行系统互联，提供性能、价格/性能和可伸缩性间的最佳组合。

## 2.2 集群节点

虽然 HP 的基于 IPF 的所有服务器和工作站都支持集群功能，但 HP 的基于 IPF 集群系统产品采用基于 Itanium2 的 Integrity 系列入口级服务器和工作站作为集群节点，包括：

- 1 路 HP 工作站 zx2000
- 2 路 HP 工作站 zx6000
- 2 路 HP Integrity rx2600 服务器
- 4 路 HP Integrity rx5670 和 rx4640 服务器

这些系统可以作为集群的计算节点、也可以作为类似于 Beowulf 集群的登录和管理节点，提供作为超级计算机系统基础节点所需的浮点和整数处理能力、高速缓存容量、高速缓存的带宽、总内存容量、内存带宽、系统互联网络通信吞吐能力以及 I/O 能力。

表 7 和表 8 分别列出上述产品的基本概貌。

表 4 常用互联模式比较表

联接架构	优点	缺点	说明
平布总线	结构简单 成本低	总线上节点争用带宽，节点数不能太多	总线架构显然不适合应用于高端的集群系统
交叉交换器	提供任何两个端口之间同时直接联接，通信效率高	内部联线的个数，随端口平方增加。端口数不能太大	最佳的端口数随技术水平提高而增加；8 个端口是目前较好的选择
树结构	能够支持较多的节点	通信带宽不能随着节点数增加而扩大	在这种结构中，越向上节点越少、带宽越窄，往往在根部形成瓶颈

表 5 交叉交换架构下端口与内部互联线路的关系

外部联接端口数	交换器内部互联线路数
4	16
8	64
10	100
16	256
32	1024

表 6 混合互联架构常用选择

互联模型	说明
总线+交叉交换	成本较低，但底层仍然存在争用带宽的问题。
交叉交换+交叉交换	提供最高的带宽，但扩展余地较小 HP 的
多层胖树结构(利用交换交换器作为树结构的节点构成多层次胖树结构)	提供最高的可伸缩性，且成本比采用两层交叉交换模式低，HP 的基于 IPF 和 Alpha 的超级计算机都采用这种架构

表 7 HP 基于 Itanium2 工作站产品概貌

	zx2000	zx6000
处理器	1 路 Intel Itanium2 处理器最高主频 1.40 GHz	1-2 路 Intel Itanium2 处理器最高主频 1.50 GHz
芯片组	zx1	zx1
内存	最大 8GB DDR SDRAM	最大 24 GB DDR SDRAM
带宽	6.4 GB/s 系统； 4.2 GB/s 内存； 2.8 GB/s I/O	6.4 GB/s 系统； 8.5 GB/s 内存； 3.3 GB/s I/O
PCI-X/PCI 槽	5 PCI-X, 1 4X AGP	3 PCI-X, 1 4X AGP
港湾数	内部-2, 外部-2	内部-3, 外部-1
内部存储	最大 282 GB	最大 438 GB
端口	LAN-1, USB-4, series-2	LAN-2, USB-4, series-2
操作系统	HP-UX 11i v1.6, Windows, Linux	HP-UX 11i v1.6, Windows, Linux

表 8 HP 基于 Itanium2 服务器产品概貌

	rx2600	rx4640	rx5670
处理器个数类型	1 或 2 个主频为 1.3GHz 或 1.5GHz Itanium2 处理器	1,2,3 或 4 个主频为 1.3GHz 或 1.5GHz Itanium2 处理器	1,2,3 或 4 个主频为 1.3GHz 或 1.5GHz Itanium2 处理器
内存容量	1-24 GB	1-64 GB	1-96 GB
内存带宽	8.5 GB/s	12.8 GB/s	12.8 GB/s
芯片组	HP zx1 芯片组		

扩展插槽	1 个 PCI-X 插槽, 1 GB/s 持续带宽, 64-位 133MHz 3 个 PCI-X 插槽, 0.5GB/s 持续带宽, 64-位 133MHz 每个插槽都是全长度的, 有独立的总线	2 个 PCI-X 插槽, 在独立总线上, 1 GB/s 持续带宽 64-位 133MHz 4 个 PCI-X 插槽, 在 2 条共享总线上, 0.5GB/s 持续带宽, 64-位 66MHz, 0.5GB/s 持续带宽, 6 个 PCI-X 插槽, 在 3 条共享总线上, 64-位 66MHz, 0.5GB/s 持续带宽, 1 个 PCI 插槽用于图形/USB, 64-位 66 MHz, 0.5GB/s 持续带宽	3 个 PCI-X 插槽, 在独立总线上, 64-位 133MHz, 1 GB/s 持续带宽 6 个 PCI-X 插槽, 在 3 条共享总线上, 64-位 66MHz, 0.5GB/s 持续带宽, 1 个 PCI 插槽用于图形/USB, 64-位 66 MHz, 0.5GB/s 持续带宽
热插拔磁盘驱动器	3 个港湾, 3.5 英寸磁盘 438 GB 最大内部存储集成的双通道 Ultra320 SCSI 控制器	2 个港湾, 3.5 英寸磁盘 292 GB 最大内部存储集成的双通道 Ultra160 SCSI 控制器	4 个港湾, 3.5 英寸磁盘 584 GB 最大内部存储集成的双通道 Ultra160 SCSI 控制器
可移动介质	1 个滑轨介质港湾, 可选择 16X -ROM 或 16X/10X/40X CD-RW	可选择 DVD-ROM 或 DVD(与 CD 写兼容)	1 个供可选 SCSI 设备使用的港湾, 可以选择 DVD-ROM 或 DDS-3
核心互联端口	Gigabit-TX LAN , 10/100 BT LAN , Ultra320 SCSI, 2 个通用的 RS-232 串行端口, VGA, 4 个 USB 端口	Gigabit-TX LAN, 10/100 BT LAN, Ultra160 SCSI, VGA, 2 个 USB 端口	Gigabit-TX LAN , 10/100 BT LAN, Ultra160 SCSI, VGA, 2 个 USB 端口
管理处理器互联	10/100 BT 管理 LAN (Web 控制台访问)RS-232 本地控制台, RS-232 远程/Modem 控制台, RS-232 通用		
高度	2 U	4 U	7 U

这些产品领先的基准测试和应用性能、性价比、I/O 和网络通信功能、高可用性和支持多操作系统等特性为在它们基础上建立集群架构的高端和超级计算机系统提供了优异的条件(详见[15])。

## 2.3 互联设备

HP 基于 IPF 的超级计算机采用多层次树状互联架构, 利用当前最流行、最领先的 Myricom 公司的 Myrinet 和 Quadrics 公司的 QsNet 作为内部互联网络、联接装备 IPF 的服务器或工作站, 建立提供高性能和最高性价比的集群系统。

### 2.3.1 Myrinet 互联网络

Myrinet 是 Myricom 公司推出的高带宽、低延迟、无阻塞分组交换网络。Myricom 的目标是把商品化产品作系统互联构成计算机集群。Myrinet 是当前应用最广的集群系统和超级计算机内部互联网络。根据 2002 年 11 月公布的 TOP500 清单, 世界上最大的 500 台超级计算机中有 28%, 即 140 台采用 Myrinet 技术构成互联网络。HP 在 Myrinet 技术领域具有明显的领先地位。事实上, HP 用于构建 UNIX 集群的 HyperFabric 和 Hyperplex 互联网

络，本身就是 HP 品牌的 Myrinet 网络。HP 利用 Linux 集群构建的基于 IPF 超级计算机也使用 Myrinet 作为内部互联网络。

## Myrinet 互联网络架构

Myrinet 是一个高带宽、低延迟、无阻塞互联网络。为了实现高性能和低成本，Myrinet 采用基于 16 端口交换器的多层胖树互联架构。

Myrinet 的基础部件 是 16 端口交叉交换器，提供联接 8 个主机和 8 个联接上层网络的端口。端口与端口以交叉交换方式实现无阻塞通信，延迟不超过 300ns。

## Myrinet 互联网络设备

Myrinet 互联网络设备包括 Myrinet 交换器、Myrinet 主机接口和相应的通信软件。

### Myrinet 交换器

Myricom 公司当前提供 Myrinet-2000 交换器产品系列，包括 8、16、32、64、128-端口的交换器，双向带宽达到 10 GB 以上、延迟不超过 300 ns、功耗为 6-11 w。

Myrinet 交换器中使用分块切入的分组路由。多端口交换器通过联线与其他交换器或者任何网络中的单端口主机接口相连接。每个交换器内部有流水线的交叉交换器，带有流控制和输入缓冲区。Myrinet 分组是任意长的，它可以携带任何类型的分组而不需要适配层。

### Myrinet 主机接口

Myrinet 主机接口应用于将集群节点联接到互联网络上。Myrinet 主机接口基于一个 32b 定制 VLSI 处理器（称为 LANai 芯片）、带有 Myrinet 接口、分组交换接口、DMA 引擎和快速 SRAM。SRAM 用于存储 Myrinet 控制程序(MCP)以及分组缓冲区。这一微架构在节点的通用总线和 Myrinet 网络联线之间提供一个灵活和高速的接口。Myricom 目前销售适用于 PCI 和光纤通道的 Myrinet 主机接口。

### Myrinet 控制软件

Myrinet 控制软件在接口处理器上执行，以减少操作系统开销，但驱动程序和操作系统仍

然在主机上执行。Myricom 提供标准的 TCP/IP 和 UDP/IP 接口以及 Myrinet API。

## Myrinet 的可伸缩性

Myrinet 具有很高的可伸缩性，能够扩展到支持规模非常大的集群系统。图 8 表示一个支持 512 个节点的 Myrinet 互联网络。该网络采用基于 160 个 16 端口交换器的 3 层胖树架构(总共拥有 2560 个交换器端口)，能够支持由 512 个节点组成的集群系统，提供 1.024 Terabits/s (128 GigaBytes/s) 的聚合带宽。类似架构的 Myrinet 互联网络可以扩展到支持拥有 8192 个节点的集群系统。

### 2.3.2 QsNet 互联网络

HP 也支持用户利用 Quadrics 公司著名的 QsNet 网络产品来建立基于 IPF 的超级计算机系统。QsNet 支持超级计算机系统内部互联的基础设备由安装在一个 QM-S16 16 端口低轮廓独立机箱或一个 QM-S128 128 端口可伸缩交换器机架中的网络交换器卡组成。网络交换器卡使用交叉交换技术提供点到点的联接、可伸缩带宽和低延迟。

网络交换器的基本构件是 Quadrics 交叉交换芯片。这一芯片是一个用来构成联接所有节点的多级网络联线的 8 路交叉交换器。这一交换配置允许同时联接所有节点。

每个适配器为节点提供与由 8 路交叉交换器构成的胖树网络的联接路径。

数据网络可以重复，以增加带宽(使用节点的多个独立 I/O 总线)和容错能力。每个节点可以拥有二个或多个网络适配器，把它们联接到不同的网络层(有时称为轨)。

## QM-S128 128 端口交换器架构

QM-S128 128 端口交换器内部采用基于 8 端口交叉交换器的 3 层胖树架构。

QM-S128 128 端口交换器采用 8 端口交叉交换器作为联接节点。所有联接节点从上到下分层排列。最低层交换器用一半端口联接计算机系统中的计算节点、另一半端口联接上层节点；上面各层的交换器的端口一半用于联接上层交换器、一半用于联接下层交换器；顶层的交换器的 8 个端口全部用于联接下层交换器。

## QsNet 互联网络组成

QsNet 互联网络由三个硬件部件组成：插入系统节点的 Elan 适配器卡、16 端口或 128 端口互联交换器和超高频互联电缆。

## Elan 适配器卡

Elan 适配器卡是一个基于 Quadrics Elan 通信设备的高性能网络接口卡。超级计算机系统中每个节点需要一个 Elan 适配器卡。该卡提供与系统高速网络交换器(16 端口交换器或 128 端口交换器)联接的高速接口。目前的型号是 Elan3，以后将生产性能更高的 Elan 4。

Elan 适配器卡是一个智能设备，内部实施通信协议。它使用一个硬件 DMA 接口，能够以极少的系统开销传输数据。测试的结果表明：通过 Elan 适配器卡进行 MPI 通信，只占用 2% 的 CPU 时间，而通过千兆位以太网进行 MPI 通信却需要占用 35% 的 CPU 时间。Elan 适配卡还留有足够的性能余地支持未来更新的系统。

每个基于 IPF 系统的多 PCI 总线设计，使它能够有效地支持两块 Elan 适配卡和另一个高性能设备如光纤通道存储设备。在 64 位/66 MHz PCI 插槽上，一个适配卡能够提供 280 MB/s 的通信能力，两块适配卡能够提供 500 MB/s 的通信能力，满足系统内部通信的需要。

## QM-S16 16 端口交换器卡

QM-S16 16 端口交换器卡是一个由 8 个 Quadrics 交叉交换部件组成的网络交换器卡，它构成一个 2 级胖树，产生一个 16 路交换器，允许联接 16 个计算节点。

## 高层交换器卡

高层交换器卡是包含 1 Quadrics 交叉交换部件的网络交换器卡。它提供 128 端口交换器顶层的 8 个端口。它的 8 条联线联接来自每个 QM-S16 16 端口交换卡前端卡的给定联线。

## QM-S16 16 端口交换器

QM-S16 16 端口交换器是一个独立的高速交换设备。它提供能够联接 16 个独立计算机系统的 Quadrics 联接端口，每一个端口都与一块 Elan 适配卡相连接。该设备包含 1 个 QM-S16 16 端口交换器卡。QM-S16 16 端口交换器卡使用 8 个 Quadrics 交叉交换部件。这 8 个交换器按照两层胖树拓扑来排列，形成一个 16 路的交换器。这一 16 端口交换器 2U 高，可以安装在低轮廓独立机箱或 19 英寸标准机架中。

## QM-S128 128 端口交换器

QM-S128 128 端口交换器是一个高速的 128 端口交换设备。它提供能够联接 128 个

独立计算机系统的 Quadrics 联接端口，每一个端口都与一块 Elan 适配卡相连接。

其中的交叉交换器按照三层胖树拓朴来排列。该设备包含 8 个 QM-S16 16 端口交换器卡插槽。每个 16 端口交换器卡能够与 16 个独立计算机系统相连接，总共形成 128 个端口。这一 128 端口交换器 17U 高，可以安装在 19 英寸标准机架中。

## QsNet 的扩展

QsNet 具有很强的可伸缩性，支持规模更大的超级计算机系统。图 11 是一个具有 1024 个端口 QsNet 的示意图(包含 4 个 256 端口的子网)。

## HyperFabric2 互联网络

HP 也提供本公司专利的 Myrinet HyperFabric2 作为集群节点互联网络，满足高性能传输的需要。HP 的 HyperFabric2 产品是由高性能网络交换器、I/O 接口卡和通信软件组成的，为基于 IPF 处理器的 HP-UX 系统提供高速度、低延迟的解决方案，从而降低在集群内移动数据的延迟。HP 专利的超级消息传递协议(HMP)捆绑在 HyperFabric2 内，通过优化各种通信任务的处理大大提高并行和技术应用的性能。除了高性能外，HMP 还能够确保维持互联环境之间的高可用操作。

## 2.4 存储设备

随着数据量的爆炸性增长、数据价值的提高，IT 技术进入了以数据为中心的新时代。对支持高性能技术计算应用和/或高端企业应用的集群系统来说同样如此：不仅需要更大容量、更高速度的节点上本地磁盘存储，而且需要具有更高的可靠性、可伸缩性和可管理性的 NAS 和 SAN 架构的网络存储系统来支持基于集群的数据存储、数据交换和数据备份以及容错和容灾解决方案。目前大规模的 HPTC 和企业应用都需要配置 TB 以至 PB 级的存储系统，促使存储设备和技术飞跃发展、市场规模急剧扩大、在各种高端应用中占据越来越重要的地位。

HP 是存储技术和市场份额领先的厂商，HP 继承了康柏原来的企业网络存储架构(ENSA)战略、推出了 ENSAextended 战略，响亮地喊出了存储公用事业服务的口号，描绘了在未来世界中象水和电那样向企业用户提供存储服务的美好明天。HP 通过发展存储虚拟化和存储管理软件、StorageWorks 磁盘阵列、NAS 存储系统以及 SAN 存储区域网络基础设施等一系列硬软件新产品和存储解决方案，在技术、产品和市场等方面取得了全面的进展，保持和发展了 HP 在存储领域的领先地位。HP 领先的存储虚拟化和存储管理技术、全面的产品和丰富的解决方案也成为 HP 基于 IPF 集群系统取得成功的坚强支柱和重要的竞争优势。

表 9 HP 的存储软件产品

### 存储软件

一体化存储资源管理软件：

HP StorageWorks Command View

HP OpenView Storage area manager

高可用性管理软件:
HP OpenView Continuous Access Storage Appliance
HP StorageWorks virtual replicator
HP StorageWorks family of local and remote replication
HP StorageWorks secure path and auto path
数据管理软件:
HP OpenView storage data protector
HP OpenView storage media operation
虚拟化软件:
HP CASA
HP StorageWorks virtual replicator

表 10 HP 存储硬件设备			
	入门级存储	中档存储	企业级存储
网络联接存储(NAS)	HP StorageWorks NAS1200 HP StorageWorks NAS 4000	HP StorageWorks NAS 2000 SAN 和 NAS 融合解决方案	HP StorageWorks NAS 9000
在线存储设备	VA 7110, MSA 系列	VA7110, EVA3000	EVA5000, XP 系列
SAN 基础设施	Switch 2/8 BL, 2/16 Blade 2/8 Edge 2/12, 2/24 MDS 9120, 9140	Switch 2/32 Edge 2/32 MDS 9216	Switch 2/64 director 2/64, 2/140 MDS 9509, 9506
磁带 & 光学设备	Autoloaders SSL 1016 Standalone drivers	MSL 系列 SSL 系列	光学存储设备 ESL 系列

### 三、HP 基于 Itanium2 的 HP-UX 集群解决方案

随着支持高端应用的处理器从 RISC 向 EPIC 过渡, HP 推出了利用商品化的基于 IPF 的工作站和服务器 HP-UX 的计算集群 hptc/ClusterPack、提供建立高端和超级系统的新途径, 满足高端 HPTC 应用需求。此外, HP 的 hptc/ClusterPack 集群还提供如下的关键特性:

- 通过增加节点水平扩展;
- 通过使用更大的 SMP 节点垂直扩展;
- 故障隔离: 单个计算节点故障将不会造成整个集群系统停机;
- 非对称性: 在一个集群中混合和匹配不同的节点;
- 配置的灵活性: 允许使用不同的节点和互联网络;

HP 的 hptc/ClusterPack 集群由如下的主要部件构成：

- 入口节点 - 为用户提供访问集群的入口。在较小的集群中，入口节点也可以作为管理服务器使用；
- 计算节点 - 提供集群计算能力和存储容量的服务器或工作站；
- 管理服务器 - 运行集群管理软件、提供集群中所有部件单一管理点的服务器；
- 集群 LAN——通常是一个用以监视和控制所有系统部件的 Ethernet，也可以处理与文件服务器之间的通信；
- 互联网络 - 实现计算节点之间的高速互联、提供并行应用所需的消息传递和访问远程内存功能；
- 存储设备 - 包括每个计算节点的本地磁盘空间和可选的网络联接存储(NAS)。NAS 设备直接联接在 Ethernet 网上，提供可供所有计算节点共享的、更加容易安装、维护和更可靠的存储设备；
- 集群管理软件——供系统管理员和最终用户使用的 hptc/ClusterPack 集群管理软件；
- 管理处理器 (MP) - 控制系统控制台、服务器的复位和加电管理功能；

### 3.1 hptc/ClusterPack 集群的硬件

HP hptc/ClusterPack 集群的硬件主要包括计算节点、互联网络和存储设备，详见本文第二章。

### 3.2 hptc/ClusterPack 集群软件

虽然集群能够提供更高的性价比和整合资源，IT 管理人员和最终用户担心的最大问题是集群环境的可管理性。HP 利用先进的 hptc/ClusterPack 集群管理软件解决了这一问题。hptc/ClusterPack 提供单点执行集群系统管理、集群系统资源控制和监视以及分布式工作负载管理的功能。这一软件基于 HP 的企业系统管理解决方案- HP Servicecontrol Manager 和 Platform Computing 最新的产品 ClusterWare。HP-UX 还将吸收 Tru64 UNIX 的 TruCluster 集群的功能，进一步增强 ClusterPack。

HP Servicecontrol Manager 已经广泛地应用于大规模的数据中心、通过单点控制管理几百个 HP 系统。它的多系统管理功能(如成组操作和基于角色管理等)使得客户能够实现优化的 IT 资源利用效率。安全的用户身份确认特性提供提高的安全性，根据用户的选择提供远程管理的功能。ServiceControl 还能够与 HPServiceGuard 等集群软件以及其他软件集成，提供高可用性和其他管理功能。

Platform Computing 的 ClusterWare 来源于业界领先的分布式资源管理解决方案 Load Sharing Facility (LSF)，是专门为集群管理和便于系统管理员及最终用户使用设计的。该软件提供业界最强的工作负载平衡和系统管理功能，同时使得用户能够象单个服务器一样方便地管理集群系统。Clusterware 的容易集成外来软件的架构使之能够方便地容纳各种管理工具和应用软件，也允许用户方便地进行集群配置。这一特性使得 Clusterware 赢得了“提供象单个服务器一样集群系统”计算模式的美名。

hptc/ClusterPack 把 ServiceControl 和 ClusterWare 组合在一起，提供一系列领先的特性，成为最有发展前途的基于 IPF 的 UNIX 集群系统：

- 单点集群管理，确保容易管理集群实施和集中配置集群的故障恢复；
- 满足高性能计算基础设施的工作负载管理需求；

- 提供集群环境全面视图、便于管理和监控集群状态；
- 提供节点全面状态、详细查找故障；
- 使用视图提高每个节点和整个集群的资源利用效率；
- 支持集群范围作业发送和监控、以便使用；
- 详细监控集群内各种作业的运行性能；

### 3.3 基于 HP-UX 集群解决方案的应用

HP 具有提供基于 PA-RISC 处理器 HP9000 服务器的 HP-UX 集群系统的悠久历史。HP-UX 操作系统下的 ServiceGuard 集群软件已经售出了 10 多万个许可证。HP 的 Hyperplex UNIX 集群系统使用基于 PA-RISC 处理器的 SuperDome 服务器作为计算节点、使用 HP 品牌的 Myrinet-HyperFabric2 作为互联网络、运行 HP ServiceGuard 和 ServiceControl(以及 Platform Computing 公司的 LSF 和 ClusterWare)等集群管理软件，在 TOP500 中占据 127 个位置，明显领先于其他厂商。HP 的 Hyperplex 集群目前广泛应用于电信、金融、汽车制造业、电子商务、政府部门、科学研究、数据仓库等许多领域。基于 IPF 的 hptc/ClusterPack 集群是 HP 贯彻全面向 IPF 过渡战略的产物，是对原有的 Hyperplex 集群的继承和发展。hptc/ClusterPack 集群在设计上使用与 Hyperplex 集群相同的操作系统、集群管理软件和集群互联网络，提供更高的性能、性价比和更远大的发展前景，必将更加广泛地应用于 HPTC 和其他重要行业，发展 HP 在 HPTC 领域的领先地位。

## 四、HP 基于 Itanium2 的 Linux 集群解决方案

### 4.1 HP 基于 Itanium2 Linux 集群解决方案的类型

- HP 开发了四种 Linux 集群解决方案(见图 13):
- **客户组装集群系统：**对希望自己组装集群的客户，HP 提供支持 Linux 的系统部件(如服务器、机架、电缆等)、系统软件和中间件帮助他们建立集群系统，实现最佳的价格/性能；
  - **客户在 HP 集群硬件平台上选择软件组成集群：**对希望采用 HP 组装的集群硬件、自己选择软件的客户，HP 提供基于机架安装 ProLiant (或 Integrity)服务器的 LC 系列产品。LC 集群可以扩展到 128 个处理器配置。HP 已经与多个 ISV 签订共同开发和市场协议、提供硬件设备供开发、移植和测试它们的应用软件使用，最后由 HP 进行质量确认、作为 HP 推荐的解决方案；
  - **HP 集成和支持的集群解决方案：**对希望得到完整解决方案的客户，HP 提供 XC 集群系列的交钥匙系统。XC 系统是完全集成的集群系统，它使用基于 Itanium 处理器的 Integrity 服务器作为计算节点、采用基于 Linux 操作系统的标准系统软件和 Lustre 集群文件系统；
  - **HP 推荐的集群参考解决方案：**对希望开发能够解决最大规模问题的超级集群系统的客户，HP 通过公司的专业服务部门提供基于已开发系统的参考模型、广泛的设计选择空间作为客户开发定制集群系统的出发点(HP 为西北太平洋国立实验室提供的基于 Linux 集群架构的超级计算机已经进入 TOP10 行列，是参考解决方案之一)；

HP 提供的上述四种 Linux 集群解决方案可以分为两大类：一类是 HP 支持客户自行构建基于 IPF Linux 集群的解决方案，包括：客户组装系统、客户在 HP 集群硬件平台上选择软件组成集群、HP 推荐的集群参考解决方案；另一类是 HP 提供传统集群系统解决方案。这两类解决方案都采用 Beowulf 架构作为构建 Linux 集群系统总体架构。

HP 采用著名的 Beowulf 架构来构建基于 Itanium2 的 Linux 集群系统。事实上，TOP500

中有 28 套以上的系统是基于 Myrinet 互联网络的 Beowulf 集群。Beowulf 集群不是一个具体的产品，而是一个用于利用可变数量、运行 Linux 的低端计算机建立集群系统的设计思想。Beowulf 集群的目标是以比通常低得多的成本建立一个并行计算超级计算机环境。随着高端芯片逐步由 RISC 向 EPIC 架构过渡以及互联网络和 Linux 操作系统软件技术的发展，基于 IPF 的 Beowulf 集群必将在高性能计算的顶端占有越来越重要的地位、逐步发展成为建立超级计算机的重要途径之一。

## 4.2 HP 支持客户建立基于 IPF Linux 集群的解决方案

HP 提供整套解决方案支持客户按照图 所示的总体设计框架构成基于 IPF 的、Beowulf 架构 Linux 集群系统。HP 提供的解决方案由集群节点、存储设备、互联网络、Linux 操作系统、集群系统软件、管理软件和开发工具等硬软件部件以及相应的技术服务和支持组成。

### 集群节点

HP 的基于 IPF 的 Linux 集群主要使用装备 Itanium 2 的 HP Integrity 系列入口级服务器(rx2600, rx5670, rx4640)和工作站(zx2000,zx6000)作为集群节点 (详见本文第二章或 [15])。这些系统可以作为集群的计算节点、也可以作为 Beowulf 集群的登录和管理节点。用户也可以根据需要采用企业级的 SuperDome 或中档的 rx7620 和 rx8620 服务器作为计算节点，提供更高的性能。

### 存储设备

HP 领先的 StorageWorks 存储设备、NAS 和 SAN 架构网络存储系统、为高端和超级计算机系统提供了强有力的支持。

### 互联网络

互联设备是建立集群架构超级计算机的基础。HP 基于 Itanium2 的超级计算机解决方案采用多层胖树互联架构，利用当前最流行和领先的 Myricom 公司 Myrinet 和 Quadrics 公司 QsNet 作为内部互联网络、联接装备 IPF 的服务器或工作站，提供建立高性能超级计算机系统所需的高带宽和低延迟 (详见第二章)。

### Linux 操作系统

HP 提供预装针对 zx1 芯片组进行性能优化的 Linux 的服务器和工作站以及广泛范围优质的 Linux 支持服务。几乎所有主要的 Linux 操作系统都提供支持 Itanium2 的 64 位功能(包括中国著名的红旗 Linux)。HP 的 Linux Itanium 产品把 Linux 的公开性、灵活性和低成本等优势与 Itanium 系统的高性能优势结合在一起，为许多应用领域提供了最佳选择。

### 集群软件

HP 与 Linux 集群领域中领先的 ISV 合作，在提供丰富的集群软件、管理软件和开发工具，支持利用 Linux 下 Beowulf 的集群设计思想建立基于 IPF 的高端和超级计算机系统。除了提供商品化系统外，HP 也提供全面的服务，支持用户根据自己的实际需求和条件选择适当的硬件产品，建立自己的基于 IPF 高端或超级计算机系统。HP 在基于 IPF 的集群平台上提供如下 Linux 集群软件产品：

- ClusterWare - Platform Computing 公司的 ClusterWare 集群软件提供业界最强的工作负载平衡和系统管理功能，同时使得用户能象单个服务器一样管理集群系统；
- ClusterWorX - Linux NetworX 公司提供的 ClusterWorX 集群软件的主要特点是提供很强的系统状态监控功能、事件管理功能、远程访问和管理功能以及集成的磁盘克隆功能。ClusterWorX 提供非常便于使用的 GUI 接口、供用户管理集群系统；
- MSC.Linux - MSC.Linux 集群软件是专门为高性能科学和技术计算设计的集群软件，提供很高的并行处理功能和管理 Beowulf 集群所需的所有管理工具；
- Scali UniverseXE 和 ClusterEdge - Scali 为 HP 的 ProLiant 和 Itanium2 平台开发了独特的集群软件技术和产品，提供容易使用、高安全性和高可伸缩性；
- Scyld Beowulf - Scyld Computing 公司是领先的 Beowulf 和高性能集群的开发厂商。该公司的 Scyld Beowulf 集群软件被称为第二代 Beowulf 集群软件。Scyld Beowulf 软件具有简化集群集成和设置、容易管理和管理工作量最小、高可靠性和无缝集群扩展等一系列特性；

## 系统管理软件

为了管理由大量节点组成的系统资源，HP 基于 Itanium2 超级计算机系统解决方案提供先进的作业管理、资源管理和运行管理功能，提高系统资源使用效率、简化管理、降低总拥有成本，包括：作业管理、系统状态管理、负载平衡管理、配置管理等。

### 软件开发工具

HP 基于 Itanium2 高端和超级计算机系统解决方案提供齐全和优质的软件开发工具，支持用户方便和高效地开发和移植软件，包括：多种编译程序、子程序库、查错软件、性能分析和优化软件、移植工具等。

## 4.3 HP 提供的成套基于 IPF Linux 集群解决方案

HP 基于 Itanium2 的 XC 高性能计算集群解决方案是一个以新颖设计思想和先进技术、为用户提供高性能和高性价比的集群架构超级计算机的解决方案。XC 集群系统是 AlphaServer SC 后续产品，具有一系列领先特性、满足 HPTC 和高端企业应用的需求：

- **全面的高性能：**提供全面高性能，包括：能够在最短的时间内完成最大的计算的高计算能力；提供高达几十以至几百 GB 的内存容量、大磁盘存储容量、高系统带宽，满足高端应用需求；
- **高度的通用性：**提供高度通用超级计算机，满足 HPTC 和广泛类型企业应用的需要；
- **使用商品化部件：**XC 集群解决方案完全使用商品化部件，包括处理器、互联设备、I/O、存储、操作系统、编译程序、编程工具和应用软件；
- **推动 Linux 高端应用：**实现充分利用 Linux 的开放性、推动 Linux 高端应用的解决方案；
- **提供平衡的可伸缩性：**在设计中非常注重于提供平衡的可伸缩性，包括：确保性能的平

衡发展、提供最佳的兼容性和投资保护、最大的扩展空间、最高的扩展效率和最低的扩展成本；

- **更高的 RAS 特性：**提供更强的可靠性、可用性和可维护性(即 RAS 特性)，满足支持高端关键任务应用的需要；
- **更高的可管理性：**具有更高的可管理性，不仅功能强、而且允许使用与商品化系统相同的工具管理，从而提高使用效率、降低管理开支和总拥有成本；
- **齐全的开发工具：**提供各种高水平语言、并行算法程序库、调试和查错、性能分析和优化等软件工具也日益丰富和齐全，有力地促进了推广应用；

HP 的 XC 高性能计算集群解决方案有两个产品：

- HP XC6000 是工厂集成的成套集群系统，使用基于 Itanium2 的、Quadrics QsNet 互联网络和 XC 系统软件；
- HP XC3000 是工厂集成的成套集群系统，使用基于 Xeon 处理器的 ProLiant 服务器，Myrinet 2000 互联网络和 XC 系统软件；

表 11 HP 提供的 Linux 集群成套解决方案简表

	HP XC6000 集群	HP XC3000 集群
处理器	1.3 和 1.5 GHz Intel Itanium2 6M L3 缓存	Intel Xeon 3.06 GHz, 装备 533 MHz FSB
计算节点	HP Integrity rx2600	HP ProLiant DL380 G3, DL360 G3
高速互联设备	Quadrics ELAN	Myricom Myrinet XP
Linux 销售包	符合 Linux 标准基础(LSB)的 Linux 销售包，得到 HP 的全面支持	
中间件	XC 系统软件，支持单一系统映像(SSI)和高功能的 Lustre 文件系统，加上 Platform 的 LSF 和 HP 的 MLIB	
服务和支持	提供全面服务选购件，包括 CorePaqs、Bronze 支持合同、白金服务水平。专门为 XC 集群设计的咨询和集成服务，包括程序管理、培训和启动服务	

下面介绍 XC6000 集群的硬软件组成部件。

### 4.3.1 XC6000 集群硬件

XC6000 集群硬件包括集群节点和互联系统。

## **集群节点**

XC6000 使用 HP Integrity rx2600 服务器作为计算和管理节点，其基本特性见第二章。这些产品领先的基准测试和应用性能、性价比、I/O 和网络通信功能、高可用性和支持多操作系统等特性为在它们基础上建立强大的超级计算机系统提供了优异的条件。

## **互联网络**

XC6000 使用 Quadrics 公司著名的 QsNet 网络产品来建立基于集群架构的超级计算机系统。QsNet 支持超级计算机系统内部互联的基础设备由安装在一个 QM-S16 16 端口低轮廓独立机箱或一个 QM-S128 128 端口可伸缩交换器机架中的网络交换器卡组成。网络交换器卡使用交叉交换技术提供点到点的联接、可伸缩带宽和低延迟。XC6000 的节点通过 Elan 适配器卡与互联设备联接。Elan 适配器卡是一个基于 Quadrics Elan 通信设备的高性能网络接口卡。超级计算机系统中每个节点需要一个 Elan 适配器卡。该卡提供与系统高速网络交换器联接的高速接口。目前的型号是 Elan3，以后将生产性能更高的 Elan 4(详见第二章)。

### **4.3.2 XC6000 集群系统软件**

XC6000 使用 XC6000 系统软件来管理集群架构的超级计算机系统，包括 XC 专门设计的文件系统、基于 Linux 的集群管理系统和著名的 Platform Computing 公司的 LSF5.0 负载平衡软件以及丰富和领先的开发工具。

HP XC 集群系统软件是一个完整的软件环境，包括 Linux 核心和系统命令、集群管理软件、资源管理和调度以及消息传递接口(MPI)。下面简要地说明 XC6000 系统软件各个组成部分的功能。

## **Linux 操作系统**

XC 集群所使用的 Linux 操作系统与 Red Hat Enterprise Linux AS 2.1 销售包兼容，加上 XC 专门的修改以支持所需功能。HP 把提供补丁和升级软件作为标准的支持服务组成部分，支持把标准的 Linux 升级为 XC 的系统软件。

## **用户视图**

XC 集群继承了 AlphaServer SC 优点为最终用户提供硬件的单一系统映像特性，包括：

- 支持一次登录；
- 一体化的程序开发环境；
- 一体化的作业发送系统；
- 单一文件系统名字空间，提供一致性访问用户数据功能；

## **工作负载和资源管理**

有效的资源管理利用一个有效的队列管理系统，它根据场地定义的策略和优先度完成提交给系统的批量或交互作业，它也利用一个能够高效使用硬件资源的调度系统。

XC 系统提供基于 Platform LSF v5.1 的直接作业分配和策略调度的批量和交互作业队列管理功能。XC 是通过调度而不是直接分配来管理作业的。HP XC LSF 提供一组丰富的调度策略、根据静态和动态属性的组合来调度作业和确定作业的优先度。用于调度的属性包括处理器数、作业属性、时间界限和 uid, gid, account id, project id 等用户属性以及根据优先度分配的资源和时间份额。LSF 的调度策略包括先来先服务、公平分配、分层公平分配、截止时间约束等等。极其可靠的基于网络的队列机制为集群作业管理提供最大的灵活性和集中管理功能、确保完成发送给 XC 系统的所有作业。

## 配置和管理

配置工具支持系统初始化、日常配置(或重新配置)和自举。XC 系统还提供管理工具控制系统的日常运行。系统能够探测到各种节点故障，包括节点没有响应、风扇故障、温度超出范围等。系统初始化和系统软件升级两者都是并行的。系统自举过程是模块化的，能够自举单个系统部件、保持整个系统继续运行。

## 系统监控

系统监控功能包括收集、存储和读取重要的系统参数，包括 CPU 和内存利用效率、系统和 CPU 温度、风扇速度以及其他各种有助于管理员了解整个系统性能和利用效率的参数。XC 系统软件使用监控进程(dmond) 和轮流监控进程(dmonpoller) 来收集系统参数，使用 SQL 数据库来存储和管理系统参数。

## 编译程序和工具

HP XC 系统软件能够支持今天 Linux 集群上大多数开发和性能优化工具，例如：

- Intel C++ compiler for Linux
- Intel Fortran compiler for Linux
- Intel math kernel library
- Intel Vtune performance analyzer
- Intel thread checker

HP 高效的数学子程序库也已经移植到 Itanium2 平台上，可以在 XC6000 系统上使用。HP 也与 Etnus, LLC and Pallas GmbH 合作，保证用户能够在 XC 系统上使用它们的开发工具。

## Lustre 文件系统

XC 系统软件从计划在 2004 年推出的第二版中将支持美国能源部 ASCI Path Forward 项目开发的面向对象和可伸缩的新颖文件系统——Lustre 文件系统。这是一个开放源代码文件系统，具有如下的特性：

- **提供单一可共享映像**：支持单一名字空间和一致性的并行存取；
- **支持高带宽文件传输**：支持 1Gb Ethernet, 10 Gb Ethernet, Myrinet, Quadrics 互联网络和光纤通道、支持多个数据服务器下的并行文件系统、提供很高的可伸缩性；
- **可伸缩的存储**：支持可伸缩的元数据存取，磁盘容量可以扩展到 TB 级；
- **把 SAN 和 NAS 的优点结合在一起**：支持 NAS 风格的共享数据和高可伸缩性、支持类似于 SAN 的高带宽和低开销存取，具有高度的自愈能力；
- **支持多路联接**：以比光纤通道低的成本联接几百、几千客户机；

#### 4.3.3 XC6000 集群系统客户价值

HP 基于 Itanium2 的 XC 高性能计算集群解决方案提供很大客户价值，实现了原定的设计目标：

- **提供高性能和巨大的扩展空间**：XC6000 在提供领先的性能同时，还提供巨大性能扩展空间，允许用户通过使用新一代 IPF 处理器、性能更高的节点、更多的节点和更高的带宽，最大限度地扩展性能；
- **提供最高通用性**：XC6000 在重突破传统超级计算机系统仅仅面向很狭窄的应用领域的局限，提供很高的通用性和灵活性，有力地促使该系列产品向通用化方向发展、满足广泛应用领域的需要；
- **支持全面的解决方案**：HP 在许多领域的全面解决方案，都可以借助 Linux 操作系统移植到 XC6000 上，推动它向通用化方向发展；
- **采用商品化的部件**：XC6000 提供使用商品化的批量部件、商品化的通用软件和商品化的外围设备，提供最佳的价格/性能；
- **提供平衡的可伸缩性**：XC6000 在设计中十分注重确保在处理器速度提高的同时、系统其他部件的性能的同步提高，提供保持整个系统平衡发展的可伸缩性；
- **高 RAS 特性**：XC6000 顺应超级计算机技术的发展潮流，提供支持大规模 HPTC 和高端关键任务企业应用所需的 RAS 特性；
- **高可管理性**：XC6000 提供兼有单一系统和集群体系结构超级计算机系统两者特点的管理功能，提高系统运行效率，降低总拥有成本；

### 4.4 HP 基于 IPF Linux 集群解决方案的应用

2002 年 5 月 IPF 处理器系列第二代产品 Itanium2 问世后，基于 IPF 和 Linux 集群架构的高端系统和超级计算机应用日益广泛。随着 IPF 系列和 Linux 集群技术的发展，基于 IPF 的高端和超级计算机系统将在高性能技术计算最顶端占据越来越重要的地位，并进而向更加广泛的企业应用领域发展，推动 IPF 处理器成为高端应用的主流平台。

#### 最丰富的系统资源

企业级服务器必须为企业用户提供最丰富的系统资源，包括足够数量 CPU、主存储器容量、I/O 和网络联接能力，满足企业级商业应用和高性能技术计算在计算速度、信息处理

和存储、支持大量同时用户等方面的需求。表 4-6 列出各厂商主要企业级服务器所提供的系统资源容量，显示了 Superdome 在系统资源方面的领先地位。

表 4-5 列出了服务器的主要资源容量，它们是企业级服务器发挥其支持企业核心应用骨干作用的物质基础：

- **CPU 处理能力：**服务器的 CPU 处理能力不是决定于 CPU 的个数，而且决定于 CPU 总的处理能力。Superdome 能够扩展到支持 64 个当前性能指标最高的 1.5GHz Itanium2 处理器，能够提供最高的计算能力。事实上，装备 64 个 Itanium2 的 Superdome 服务器的许多性能指标不仅超过 CPU 数量较少的服务器，而且明显超过装备 2 倍 CPU 的 Sun 服务器和 Fujitsu 服务器；
- **内存容量：**64 位处理器最大优点之一是打破了 32 位处理器 4 GB 物理内存容量的限制，允许装备容量更大的物理内存。Superdome 不仅能够提供创记录的 512 GB(可扩展到 2TB)单一地址空间的共享物理内存空间，而且也提供很强的超大规模内存(VLM)功能，能够把大规模数据集“固定”在该服务器的物理内存内、供程序访问，而不必反复访问磁盘(相当于高速内存作为大型商业数据库高速缓存)。Superdome 的高内存容量提供高效地处理最大和要求最高的企业应用的能力，也为用户和软件开发商提供了提高软件性能的强大工具和无限商机；
- **I/O 和网络联接能力：**企业级服务器必须提供强大的支持大容量和高速磁盘、磁带、各种外设以及网络通信的能力。I/O 可扩展性越大，服务器支持企业信息中心的能力越强。新推出 Superdome 支持 192 个 PCI-X 适配器(吞吐能力是 PCI 一倍以上的)的 I/O 容量为企业级联接设置了新的标准。这样规模的互联能够提供支持现代企业大规模数据存储设备、SAN 或 NAS 体系结构的网络存储系统和网络通信所必需的资源；

## 最佳的系统基础设施

企业级服务器一方面必须拥有丰富的资源，另一方面必须建立能够充分发挥系统资源潜力的基础设施。两者缺一不可。服务器的基础设施指把处理器、内存模块、I/O 端口灯资源联接在一起、协调地操作和通信的设备和机制，包括：

- **系统体系结构：**指决定处理器访问内存方式和共享关系的系统总体设计思想；
- **互联拓扑：**指处理器、内存、I/O 设备、网络通信设备联接方式和通信路径；
- **互联设备：**指按照系统体系结构和互联拓扑设计、把各种系统资源联接成一个完整系统的设备。因此，系统互联设备是在物理上就是按照系统总体设计、组成服务器系统的基础设施；当前各厂商都使用芯片组作为系统联接到核心设备。

Superdome 利用领先的 sx1000 芯片组作为互联设备、实现两层交叉交换的互联拓扑，建立了符合服务器未来发展方向的 NUMA 体系结构企业级服务器，实现兼有 UMA 和 NUMA 两者优点、克服两者缺点的设计目标：

- **降低远程和本地访问延迟比：**互联基础设施的分层交换拓扑以及高带宽，使得远程访问延迟与本地访问延迟之比仅为 2:1 并且这一延迟比例在高系统负载下只有非常小的增长。这使得基于 UMA 体系结构下开发应用软件，可以很容易地、甚至直接移植到高端的；

- **减小访问远程内存概率：**Superdome 每个单元可装备 32GB(可扩展到 128GB)的本地内存，在 HP 领先的 HPUX 11i 操作系统调度下，能够大大减少访问远程内存的概率；其他服务器的本地内存容量都小于 Superdome，加大了访问远程内存的概率，必将影响在安装和运行大型数据库和商业应用软件的性能；

从互联拓扑来分析，Superdome 和竞争的企业级服务器采用两种不同的互联拓扑结构：

- **多层交换互联拓扑：**Superdome 采用分层交换的拓扑结构，即由两层交换器构成系统的互联基础设施。第一层是采用交叉交换器互联的单元构件，第二层是全程交换器；
- **总线-交换互联拓扑：**Sun Fire15000，IBM p690 采用总线-交换拓扑结构。第一层是总线，把若干个 CPU(或内存模块)联接在类似于总线互联设备上。第二层是交换器；

Superdome 的多层交换互联拓扑和领先的芯片组技术，使系统带宽能够随着系统扩展而增加，提供最大的聚合带宽和每个处理器带宽：

- 64 GB/s 交叉交换后面板互联带宽( $16*8\text{GB/s}/2$ )；
- 256 GB/s 内存聚合带宽 ( $16*16\text{ GB/s}$ )；
- 32 GB/s IO 带宽( $16* 2\text{GB/s}$ )；
- 204.8 GB/s CPU 聚合带宽( $64*3.2\text{ GB/s}$  每个处理器带宽)；

## 降低系统延迟和消除瓶颈

现代计算机系统设计的基本矛盾是：处理器的速度不仅比内存快得多，而且处理器速度提高的速率也比内存速度提高的速率高得多。此外，如何利用有限的 I/O 带宽满足高带宽网络通信以及磁盘、磁带等高速外设信息传输速度要求，也是一个难题。因此计算机系统如果设计得不好往往会造成内存瓶颈或 I/O 和通信瓶颈，使系统资源不能有效地发挥作用。Superdome 服务器采用先进的系统设计和领先的实施技术消除了系统瓶颈，实现了无瓶颈设计，使所有资源能够充分发挥作用，并留有足够的发展余量：

- **提供最大的内存带宽：**Superdome 支持 256 路交叉的 256GB 聚合内存带宽，使得内存不会成为系统瓶颈；
- **不变的每个处理器带宽：**Superdome 在系统单元数从 4 个增加到 16 个时或者系统负载增加过程中始终保持每个处理器平均享受

3.2 GB/s 的带宽(4 个处理器共享 12.8 GB/s 的带宽)，使处理器不会因为不能及时取出指令和数据而被阻塞，不能发挥作用；

- **创新的高速缓存一致性协议：**为了加快内存的工作速度，必须采用高速缓存，由此也带来了高速缓存一致性问题。Superdome 采用全新的、无等待、低占用率的高速缓存一致性协议，避免了当前常用一致性协议的分块访问和重发访问的问题，大大降低了系统的延迟；
- **新颖的 CC-NUMA 技术：**在多处理器系统中，每个内存模块都有自己的高速缓存。因此，高速缓存是分布式的。过去，一般通常采用基于侦听的一致性机制(snoop)来解决分布式缓存一致性问题。Sun Fire15000 和 NEC 也采用这种机制。基于 snoop 机制的一致性

往往会产生巨大的单方面延迟时间和带宽不足，从而导致出现系统瓶颈。Superdome 服务器对传统的 CC-NUMA 技术作了较大的发展、采用高度优化的目录一致性设计，占用空间小、易于执行、成本低，不对系统带宽带来任何限制。它能够以非常低的一致性带宽管理开销，扩展到 16 个单元；

- **分布式 I/O 带宽和连接：** Superdome 分布在 16 个单元上 192 个 I/O 卡，每个卡都有自己专门的 I/O 总线；在 PCI-X 适配器级，聚合带宽可以达到 128 GB/s；系统核心(处理器和内存)与 I/O 控制器之间的可用 I/O 带宽可达 32 GB/s。在合理的安排下，不会形成 I/O 和网络通信瓶颈，满足了企业级服务器支持高速磁盘和网络存储系统以及高带宽网络通信的需要；
- **平衡的信息传输和通信：** Superdome 实现了点到点全局“信息包”交换的通讯机制，使得处理器、内存和 I/O 之间信息流量保持很好的平衡，不会形成阻塞和瓶颈；  
模块化设计和可伸缩性

Superdome 采用先进的模块化设计技术。系统的基础模块化构件是单元模块。每个单元提供最大的资源容量：

- 最多提供 4 个 1.5 GHz 的 Itanium2 处理器模块；
- 在最多 4 个承载部件上提供最多可达 32GB 的内存存储器；
- 支持多达 12 PCI-X 插槽；
- 8 GB/s 互联带宽；
- 16.0 GB/s 聚合内存带宽；
- 2.0 GB/s I/O 带宽；
- 3.2 GB/s 的每处理器带宽 (4 个处理器共享 12.8 GB/s)；

HP Integrity Superdome 服务器为操作系统提供对称多处理 (SMP) 编程模型，使任何处理器都能够访问系统上任何地方任何字节的内存。事实上，它是第一个利用关键任务 UNIX 系统来访问分布式处理器和内存的，原因如下：

- 可用带宽能随系统规模而扩展；
- 真正的系统资源硬件隔离——处理器、内存和 I/O 资源相互真正隔离，以提供系统在使用上的灵活性，这意味着：系统具有极高的 SMP 性能能够完成大量的单工作负载；资源的硬件强制隔离与出色的单系统高可用性及可管理性相得益彰，实现强大的整合平台。

### HP 独特的 sx1000 芯片组和 mx2 扩展模块

当前许多公司如 HP、Intel、日立、IBM 和 NEC 等都提供支持安腾 2 的芯片组。其中，只有 HP 的 sx1000 芯片组和 mx2 扩展模块相结合能够支持 128 个 Itanium2 处理器，提供最高的可伸缩性。

HP 的 sx1000 芯片组能够支持 8, 16 路 Integrity 中档服务器、16, 32 和 64 路的 Superdome 服务器。与其他公司的高端芯片组相比较，在设计、内存带宽和延迟等方面有明显的优势，为 Integrity 中高档服务器确立性能和性价比优势提供了坚实的基础。

#### 4.4.3 Superdome 的性能领先优势

HP Integrity 企业级服务器提供全面领先的基准和应用测试指标,以最高的性能合性价比满足企业用户的需求。

## 全面领先的应用性能

HP 在工业标准平台上的技术优势确保其基于 Itanium2 入口级服务器在主要的企业应用领域提供全面领先于竞争对手同档次的系统,有力地促进了 Itanium 向广泛的市场领域迅速扩展。表 4-8 说明基于 1.5 GHz Itanium2 的 Superdome 服务器已经成为性能最高的企业级服务器,具有全面的领先地位。

### 领先的 tpmC 指标

TPC-C 测试给出系统每分钟进行交易数量 tpmC 指标,是企业用户测量系统在线事务处理(OLTP)和支持同时用户数能力最常用的指标。HP Integrity 系列 Superdome 服务器提供领先的 tpmC 指标,当前领先的前三名的 tpmC 指标都是在 Superdome 上创造的。

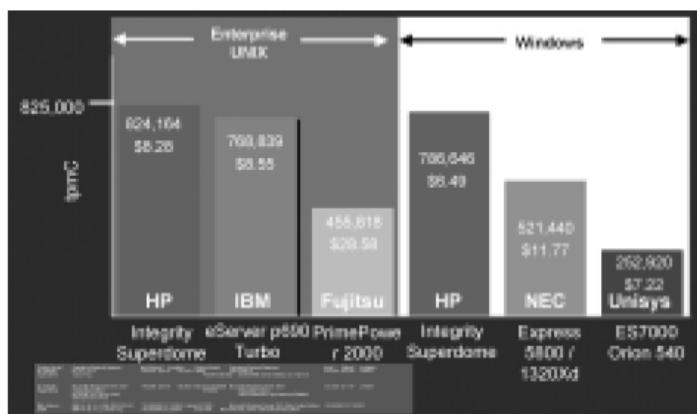


图 4-5 HP Integrity Superdome 提供 UNIX 和 Windows 下领先的 tpmc 指标

图 4-5 说明 HP 基于 Itanium2 的 Integrity Superdome 服务器在 UNIX 和 Windows 下的 TPC-C 测试指标超过 IBM 和 Sun 64 位 RISC 服务器以及 NEC 和 Unisys 基于 Itanium2 的高端服务器。

### 领先的 SPEC Rate 基准测试指标

SPECfp\_rate2000 和 SPECint\_rate2000 是测试计算机系统多处理器浮点和整数计算能力的重要基准测试指标。这一指标对于测试高端服务器的硬软件系统支持多处理器特性的具有重要意义。企业级服务器能够装备多达 32, 64 个处理器,因此 SPEC Rate 基准测试指标成为考察企业级服务器在设计和实施上是否能够充分发挥多处理器潜力、是否具有在企业计算中心或数据中心支持大量浮点和整数计算密集任务的能力、以至是否适合于在数据中心应用

的关键指标。由表 4-10 中的数据可见，基于 Itanium2 的 Integrity 系列企业级服务器提供领先于 64 位 RISC 服务器浮点和整数多处理器计算能力，体现了 IPF 性能和 HP 领先技术的优势。

#### 领先的数据仓库性能

企业级服务器经常被应用于支持数据仓库应用。TPC-H 使一个用于测试服务器系统应用于不同规模数据仓库时的性能，是量度服务器应用于数据仓库领域性能的最常用的指标。图 4-6 说明 Superdome 应用于规模为 3TB 的数据仓库时提供领先的 TPC-H 测试指标。

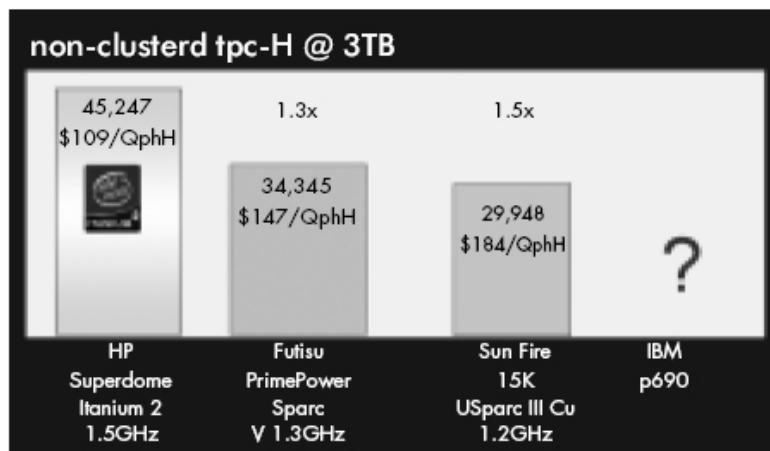


图 4-6 HP Integrity Superdome 提供领先的支持数据仓库的性能

#### 领先的 SPECjbb2000 基准测试指标

SPECjbb2000 基准测试指标用来测试服务器一侧的 Java 性能，提供测量服务器运行 J2EE (Java2 企业版)能力最客观和代表性的基准测试指标。为了能够在网对客户机一侧提供最佳的 J2EE 空间下的 Web Services，企业级服务器必须具有很高的 SPECjbb2000 基准测试指标。表 4-11 说明装备 32 个 CPU 的 Superdome 运行 J2EE 能力不仅超过装备同样数量 CPU 的 IBM p390，而且大大超过装备 72 个 CPU 的 Sun Fire 15000，可见 Superdome 特别适合于在企业网或 Internet 网上为大量客户机提供 Web Services。

#### 提供领先的平衡性能

如上所述，当前企业应用与高性能技术计算应用正在走向融合。企业级服务器处于企业数据中心，用于支持广泛类型的应用。因此，企业级服务器不仅需要具有很高的支持各种典型的企业应用性能指标(如 TPC-C、TPC-H、SPECjbb2000 等)，而且需要具有很高的支持浮点运算的性能指标(如 Linpack、SPEC Rate 等)。提供平衡的高性能已经成为考察企业级服务器主要的指标，为此要求企业级服务器在各种类型的应用中具有全面的高指标，而不是仅仅是某项(或某几项)指标特别高，而其他指标又非常低。实践的经验表明，性能畸形发展服务器往往不完全适合作为企业数据中心服务器，甚至有可能是厂商采用了专门的硬软件设计或者特殊的手段片面提高了这些指标，而在其他应用和测试中性能很可能很差。图 4-6 说明

Superdome 提供全面领先于 IBM p690 的高性能, 图 4-7 说明尽管 IBM p690 的 TPC-C 指标很高, 但其他性能很低, 而 Superdome 则由于设计上优势、能够提供全面的高性能。

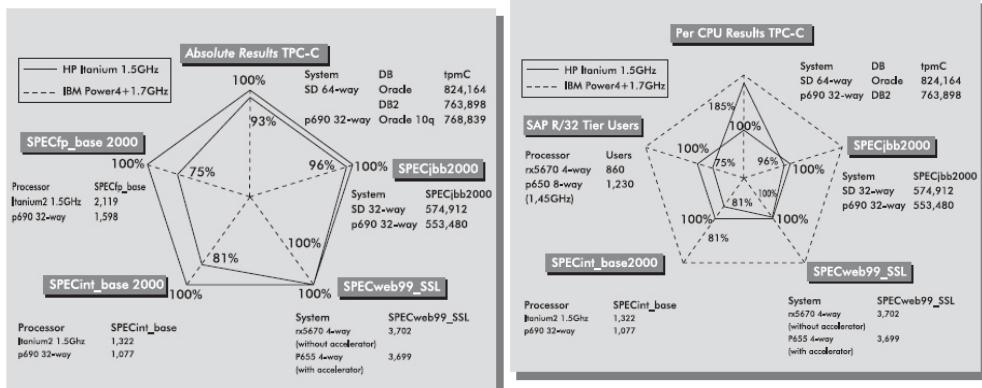


图4-6 HP Integrity在多种应用中性能全面超过IBM基于Power4+的服务器产品

图4-7 从每个CPU性能表明IBM针对TPC-C测试进行了优化、其性能是不平衡的,而HP Integrity在多种应用中性能是平衡的、能够全面满足应用需求

表 4-6 Superdome 实现最小的访问本地和远程内存延迟比		
单元板数量	CPU 数量	平均的空闲加载到使用内存延迟时间
1 个	4 个	246 ns
2 个	8 个	330 ns
4 个	16 个	371 ns
8 个	32 个	417 ns
16 个	64 个	440 ns

表 4-7 HP sx1000 与其他厂商高端芯片组的比较				
	HP sx1000 芯片组	NEC Express 5800/1000 芯片组	IBM EXA	Intel E8870
适用范围	支持基于 Itanium2 和 PA-8800 中高档服务器蜂窝状(分层交叉交换)互联体系结构	支持基于 Itanium2 中高档服务器蜂(分层交叉交换)互联系统结构, 只支持 Itanium 一种处理器	中档服务器, EXA2 是 EXA 的扩展, 适用于 Itanium	入门级服务器
处理器	8-128 Itanium2 或 PA-8800	8-32 个 Itanium2	4-16 Xeon 或 Itanium2	1-4 个 Itanium2
CPU 带宽	每个单元 12.8 GB/s	每个单元 6.4 GB/s	6.4 GB/s	6.4 GB/s
内存带宽	每个单元 16 GB/s	每个单元 12.8 GB/s	6.4 GB/s (在 x450 上 9.2 GB/s)	6.4 GB/s
I/O 带宽	每个 I/O 板 8.4 GB/s PCI-X	?	4 GB/s PCI-X	4 GB/s PCI-X
内存延迟 (传输速度)	低延迟(高速度)	高延迟(低速度)	较高(慢), 比 zx1 慢 4-9 倍	由于进行 RD 到 DDR 转换, 延迟很高

最大内存 ( 使用 2GB DIMM)	128 GB/单元 2 TB/系统	64 GB/单元, 再加 上 64GB 在子板上	64 GB (在今天的 x450 上只有 40 GB)	128 GB
其他	使用基于目录的缓存一致性机制降低开销, 单元板可以插入今天的 Superdome 系统中	使用基于 snoop 的缓存一致性机制开销很大, 影响了支持 NUMA 体系结构系统的性能	可选的 L4 缓存和内存缓冲区, 基于目录的缓存一致性	使用 snoop 缓存一致性机制, 开销很大

表 4-8 Superdome 在多种操作系统下提供全面领先的性能					
测试指标	系统	操作系统	指标	排名	
SPECint2000_rate	Superdome 8 CPU 16 CPU 32 CPU 64 CPU	HP-UX	117	117 8-路 #1	
			229	229 16-路 #1	
			453	453 32-路 #1	
			904	904 64-路 #1	
			142	超过 IBM 和 Sun 企业级服务器的最佳指标	
SPECfp2000_rate	Superdome 8 CPU 16 CPU 32 CPU 64 CPU	HP-UX	236		
			470		
			928		
			1,008,144(\$8.33/tpmC)		
			824,164(\$8.28/tpmC)		
TPC-C	Superdome Superdome Superdome	HP-UX/Oracle 10g 企业版 HP-UX/Oracle 10g 企业版 Windows/SQL	786,646(\$6.49/tpmC)	64 路 #1	
			1,008,144(\$8.33/tpmC)	64 路 #1	
			824,164(\$8.28/tpmC)	64 路 #1	
			786,646(\$6.49/tpmC)		
			1,008,604		
SPECjbb2000-java	Superdome 8 CPU 16 CPU 32 CPU 64 CPU	HP-UX	181,369	8-路 #1	
			322,604	16-路 #1	
			574,912	32-路 #1	
			1,008,604	64-路 #1	
			335(持续速度高达 1/3 TFLOPS)	超过 IBM 和 Sun 企业级服务器的最佳指标	
Linpack N*N					

表 4-9 HP Integrity 系列 Superdome 服务器的 tpmC 指标占据前三名地位					
产品	排名	记录创造日期	tpmC	\$/tpmC	操作系统和数据库软件
Superdome 64 个主频为	1	04-Nov-2003	1,008,144	8.33	HP-UX 11i, Oracle 10g
	2	30-Jul-2003	824,164	8.28	HP-UX 11i, Oracle 10g

1.5GHz 的 Itanium2 处理器	3	26-Aug-2003	786,646	6.49	Windows/SQL
-----------------------	---	-------------	---------	------	-------------

表 4-10 HP Integrity Superdome 服务器提供领先的浮点和整数计算能力						
厂商和服务 器系列	HP Superdome	HP Superdome	SunFire	SunFire	IBM eServer PSeries	Fujitsu PRIMEPOWER
处理器	Itanium2	Itanium2	UltraSparc III	UltraSparc III	Power4+	SPARC64
SPECint2000	64 路 904	64 路 453	15K 72 路 478	12K 36 路 260	--	1500 32 路 205
SPECfp2000	64 路 928	32 路 470	15K 72 路 492	12K 36 路 579	690 32 路 350	1500 32 路 346

表 4-11 Superdome 提供领先的 SPECjbb2000 基准测试指标			
	32/64 路 HP Integrity Superdome 服务器	32 路 IBM eServer pSeries 690 Turbo	72 路 Sun Fire 15000
SPECjbb2000 指标	574,912/1,008,604	553,480	433,186
CPU 类型	1.5 GHz Itanium2	1.45 GHz Power4+	1.2 GHz UltraSparc III
操作系统	HP-UX 11i V2	AIX	Solaris

# HP创建动成长企业

中国惠普有限公司

北京市朝阳区建国路 112 号中国惠普大厦

电话: 010-65643888

传真: 010-65643999

邮编: 100022

欲查询更多相关信息, 请访问 HP 网站:

<http://www.hp.com.cn>

中国惠普客户互动中心: 800-820-2255

售后服务支持热线: 800-810-5959

010-68687980



i n v e n t

最终解释权归中国惠普有限公司所有  
印制日期:2004 年 2 月北京印刷