安腾技术文集(修订版)

solution center

invent

基于 IPF 的 HP 集群系统及其应用

HP 在发展基于 IPF 集群过程中，吸取了集群技术发展几十年的经验和教训、更加坚定地走基于开放性工业标准的道路、充分地利用 IPF 处理器优势和计算机系统及网络通信技术的最新成果，为用户提供一系列先进特性和独特优点，成为推动 IPF 登上高端应用主流平台又一重大举措。

一、基于 IPF 的 HP 集群系统概述

根据当前高端应用需求、市场和技术发展趋势，HP 把支持基于 IPF 处理器服务器或工作站的集群作为高端和超级计算机系统主要发展方向，为用户提供全面的 Linux 集群解决方案和 HP-UX 集群解决。用户可以选择在 HP 支持下自行组装系统,或者采用 HP 提供的成套 Linux 集群产品，从而以最大灵活性满足自己的应用需求、计划进度和经费预算。2003 年底，HP 又推出基于 IPF 的 OpenVMS 集群，使企业用户能够在工业标准的平台上享受到 OpenVMS 集群的领先技术和独特优势。

1.1 高端和超级计算机系统的发展趋势

高端和超级计算机系统用以解决高性能技术计算应用领域中规模最大、最复杂的尖端问题。此类系统最初使用专门的技术和部件、价格非常昂贵、台数很少、市场面和应用面也很窄。20 世纪 90 年代中后期以来，许多应用领域越来越多地希望利用半导体和计算机技术发展的新成果通过更大规模、更精确的数值模拟和数字计算来进行新产品设计、提高厂商市场竞争力。许多新型的企业应用也要求使用超级计算机系统来支持更大规模的应用、存储和处理数量爆炸性增长的数据。

于是,一方面基于传统技术、昂贵的超级计算机再也不能满足日益增长的需求；另一方面，半导体(特别是 64 位处理器芯片)和计算机技术的发展又为新型的超级计算机准备了充分的条件，促使超级计算机设计思想和技术路线发生了一系列根本性的变化：人们不仅需要性能更高、满足高性能技术计算需求的超级计算机，而且需要更加通用、基于商品化部件和开放系统软件、价廉物美的新型超级计算机。它们基于新颖的硬软件系统技术，提供平衡的可伸缩性、高 RAS 和可管理性、强大的开发工具和丰富的应用软件。它们将走出高贵的殿堂和狭隘的应用领域，面向大众和更加广阔的应用领域。

为了满足人们以更低成本生产性能更高、数量更多、更加通用的超级计算机系统的需要，HP 和 Intel 联合开发的 Itanium 处理器系列应运而生。HP 基于 IPF 集群的设计思想为高端和超级计算机技术开辟新的道路,以满足不断增长的应用需求

更高、更全面的性能要求

高端和超级计算机系统主要应用于高性能技术计算和大规模的企业应用。20 世纪末，这两方面的应用都对计算机系统的性能提出来更高、更全面的要求。

高性能技术计算是利用数值模拟和数字技术方法探索和预测未知世界的技术。这一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、全球性长期气象、天气和灾害预报、工艺过程改进和环境保护等许多领域。近年来，随着研究的深入和竞争的加剧，各个领域越来越多地使用模拟的方法来解决科研和生产中的实际问题。模拟的模型越来越大、计算的精度越来越高、对超级计算机性能要求也越来越高。例如，在一个 3 维模型中，如果把从每个方向取 100 个分点增加取到 1000 个分点，对计算

机资源的需求将增加 1000 倍以上。

在基于 Internet 的新型企业应用领域，IT 系统开始从运行单一应用发展到运行复杂的应用(如 ERP 和 SCM 应用)、从面向本地发展到面向全球竞争的应用、从单一媒体发展到多媒体和流媒体处理、从手工输入发展到自动数据获取、从在线事务处理(OLTP)发展到在线分析处理(OLAP)和商务智能(BI)、从信息查询发展到辅助决策支持。由此产生的许多实际应用促使对性能更高超级计算机系统的需求急剧增长。

新型的应用对计算机性能要求主要表现在三个方面：

- **计算能力**：为了在最短的时间内完成最大的计算量，不仅需要处理能力更强的处理器 (特别是 64 位以上高精度浮点计算能力)，而且需要利用集群或大规模并行处理(MPP)架构等系统技术、支持更多数量处理器的并行计算机系统；
- **储存容量**：一个每方向 100 节点的 3 维模型需要 32 MB 内存，而每方向 1000 节点的 3 维模型就需要 32 GB 的内存。为了提高性能，往往需要利用超大规模内存(VLM)技术把整个数组放在内存中，这就需要高达几十以至几百 GB 的内存容量。内存容量增加显然也要求系统提供更大的磁盘存储容量；
- **系统带宽**：数据量的增加促使处理器和内存、内存与磁盘间的信息交换量的急剧增加。为了能够以最快的速度传输信息，要求提供足够的系统带宽，保证内存能够及时向多个处理器提供足够的数据以及内存和磁盘之间的快速和可靠的数据传输；

更高的性价比和可伸缩性

当前，不计成本的高性能计算时代已经一去不复返了。解决尖端问题的高端系统同样必须降低成本、同样要求从尽可能低入口价位开始逐步扩展，否则很难满足客户的日益紧张的经费约束和投资保护的要求。由于此类系统需要使用许多计算节点和互联设备等部件，因此必须保持每个部件的低成本。早期的超级计算机系统使用专门定制的处理器和互联设备等部件如 Thinking Machines 公司的 CM-5 和 Kendall Square 公司的 KSR-1 等，其价格非常昂贵。以后，Cray Research CRAY 公司的 T3D 和 CRAY T3E 开始使用商品化的 Alpha 处理器。当前商品化处理器和服务器性能日益提高、价格也日趋下降，为利用它们建立超级计算机系统提供了良好的基础。为此，美国政府还推出了 ASCI 计划，力图降低超级计算机系统的成本，其主要途径是尽可能采用商品化市售(COTS)硬件和软件部件，把力量集中在发展主流计算机工业不能有效地提供的专门技术。目前已经很少再有厂商使用专门的部件如向量处理器来建立超级计算机系统。今后的发展趋势是在超级计算机系统中尽可能普遍地采用商品化和大批量的工业标准部件，包括处理器、互联设备、I/O、存储、操作系统、语言、编译程序、编程工具和应用软件。人们注意到，基于开放性 IA-32 体系结构的 Xeon 和 Pentium 4 处理器的超级计算机已经在 TOP500 占有重要地位。新兴的 IPF 处理器系列必将以其开放性、大批量和 64 位寻址和处理能力，对超级计算机水平的提高产生划时代的影响，以远比 32 位开放性体系结构时代高的性能和性价比来满足日益增长的需求。HP 推出基于 IPF 的全面的集群解决方案和产品，包括 Linux 集群和 HP-UX 集群解决方案和 OpenVMS 集群系统(以及 Winodws 集群)，必将为从性能和经费两方面满足应用需求开辟新的道路，成为未来高端和超级计算机系统发展的方向。

Linux 高端应用的发展

HPTC 技术另一个新发展是 Linux 等开放源代码操作系统越来越广泛地应用于支持高端和超级计算机。20 世纪末期，支持超级计算机的主要操作系统已经逐步由厂商专门设计的专用操作系统、演变为具有较强通用性的 UNIX。这些 UNIX 系统仍然是厂商的专属产品。只有很少的超级计算机运行 Microsoft NT 操作系统和开放性的 Linux。

进入新世纪以来，Linux 操作系统得到了包括 HP 和 IBM 等最大厂商在内的几乎整个计算机产业界的支持，能够在绝大多数硬件平台上运行。64 位 Linux 也日益成熟。在新兴 IPF 平台上，Red Hat, SuSE, Caldera, TurboLinux 等许多厂商都推出了 64 位 Linux 版本，提供基于 64 位的高计算精度和巨大寻址空间。

虽然 Linux 操作系统在 PC 和桌面应用中远没有 Windows 普及，但它在服务器领域中的地位却正在飞速提高。Linux 将广泛应用于所有领域：首先是科研和技术计算等学术性较强的领域，目前已扩展到基于 Internet 的 Web 网站管理和电信等领域，今后将逐步进入所有类型的企业应用。Linux 系统技术也取得了长足的发展，已经能够提供支持高端超级计算机所需的 RAS 特性和管理功能，开发工具和应用软件也日益丰富。业内人士普遍认为：在开放源代码界的共同努力下，特别是 HP 和 IBM 等一流厂商的支持下，未来(不是现在)开放源代码的 Linux 操作系统将具有与各厂商专利的 UNIX 操作系统相比美的支持大规模 SMP 能力、RAS 特性、同样丰富的开发工具和应用软件。开放性的 IPF 的发展也将加速 Linux 的这一发展趋势。根据 Aberdeen Group 的报告，2003 年前基于 IA-32 Linux 系统仍将占据主要地位。以后，Linux 系统将逐步 64 位化，到 2005 年基于 IPF 的 Linux 系统将占据主要地位。

于是，利用 Linux 强大水平可伸缩性和大量低端商品化服务器或工作站(2-4 个处理器)建立价廉物美的"买得起"超级计算机已经成为高端应用领域中重要的新发展趋势。当前，Linux 已经在支持大型集群体系结构的超级计算机系统方面取得了一系列引人注目的成就。随着 Linux 技术和开放性的 IPF 系列的发展，人们将看到更多的基于 Linux 超级计算机系统，其应用领域也将更加广阔。

更广泛地采用集群体系结构

虽然 NUMA 体系结构的 64 位高端服务器的单系统性能和可伸缩性正在迅速提高，但是单个服务器的资源扩展上限如 CPU 个数和内存容量等仍然有较大的限制。现代高端或超级计算机主要采用以多个商品化服务器(或工作站)作为基础节点的大规模并行处理器(MPP)系统和集群系统。表 2 列出这两种体系结构的主要特点。

虽然 MPP 和集群系统技术正在相互融合，但是它们毕竟仍然是两类不同的系统。集群系统最初用于提供高可用性，后来人们又发明了能够提供超级计算能力的高性能集群。集群系统能把高性能和高可用性最佳地结合在一起，已经成为超级计算机系统的主要发展方向。特别是，随着 Linux 集群技术逐步成熟(例如 Beowulf 集群)，利用 Linux 集群构建的超级计算机系统更是与日俱增。根据 Aberdeen Group 公司的报告，Linux 集群将在今后三年内占有 80%的高性能计算市场。Linux 集群的应用也将从 HPTC 逐步扩大到广泛的企业应用领域。

与传统 MPP 体系结构相比较，利用集群体系结构特别是 Linux 集群体系结构构建超级计算机系统，不仅成本低、开放性好，且具有许多现代超级计算机所必须具备的优点：

- **支持最大规模的系统**：MPP 系统最突出的特点就是规模扩展上限非常大、能够扩展成支持拥有很多处理器和很大容量内存的并行处理系统、提供非常高的性能。Linux 操作系统

卓越的水平可伸缩性，同样能够支持最大数量的节点、构建最大规模的超级计算机系统。事实上，目前 TOP500 中越来越多的系统是具有 Linux 集群体系结构的超级计算机。这一趋势随着 IPF 处理器系列性能的提高，还将进一步发展；

- **允许使用商品化网络进行节点互联**：在 MPP 系统中，节点往往需要通过专门设计的高速网络来互联，成本很高。集群系统允许使用广泛类型的商品化网络进行内部节点互联。表 3 列出基于 Linux 的 Beowulf 允许使用互联网络。

其中，Myrinet 和 Quadrics 网络能够提供满足超级计算机内部通信需要的高带宽、低延迟和无阻塞特性(详见 2.2 节)，支持数以千计的节点。其他网络可用于支持中小规模的集群系统。

- **使用松散耦合的互联方式**：集群系统可以采用松散耦合的方式，即把网络接口与节点的 I/O 总线相联接，大大有利于把商品化的服务器或工作站直接联接到集群系统上，降低安装成本；
- **提供平衡的可伸缩性**：Linux 集群能够容纳大量独立的计算机系统，互联网络也有很大的可伸缩性，因此能够更方便地实现处理器、存储容量、系统带宽和内部通信能力的平衡、消除系统瓶颈，而且在扩展过程中能够继续保持性能的平衡、提供互相适应和匹配的高速度和精度、大容量、高带宽和通信能力，同时保持最佳的兼容性，提供安全的投资保护、最大的扩展空间、最高的扩展效率和最低扩展成本；
- **更高的 RAS 特性**：把高性能与高可用性完美地结合在一起是集群系统的最突出的优点。集群系统能够提供远比 MPP 系统高的冗余度、容错和容灾能力；
- **高的可管理性**：管理集群本来是一项十分复杂的任务。随着基于 Linux 集群技术的发展，一些厂商也能够提供类似于单一系统映象的功能以及先进的作业控制、负载平衡、分区管理、性能监控等特性，从而大大简化 Linux 集群的管理，提供较高的可管理性；

1.2 HP 基于 IPF 的产品线

根据 HP 的长期发展战略，HP 将把基于 PA-RISC、Alpha 和 MIPS 的所有服务器产品系列全部过渡到 IPF 平台上。当前，HP 主要推出 3 个层次的基于 IPF 的产品：

- **单一系统产品**：基于 Itanium2 的 Integrity 服务器系列(包括入口级、中档和企业级服务器，详见[15])和工作站(1-2 路工作站)；
- **中档产品**：8-32 节点的 HP-UX 集群系统；
- **高端产品**：32-100 (以至更多)节点基于 Linux 集群体系结构的高端和超级计算机系统；

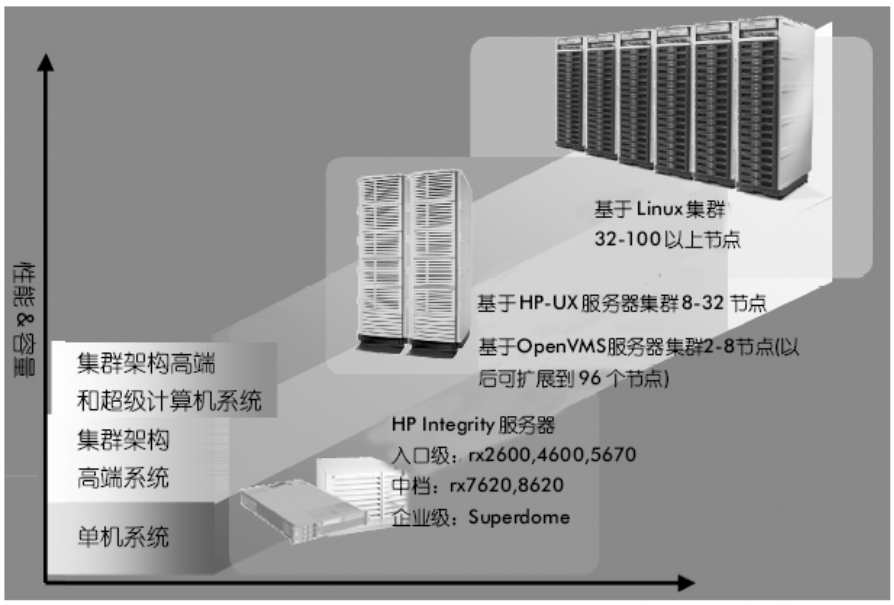

图 1 基于 IPF 的 HP 产品线

HP 这一产品策略使基于 IPF 产品以最快的速度覆盖计算机系统各个档次，推动 HP 成为第一个能够以基于 IPF 处理器产品全面满足用户各方面的需求，充分反映了 HP 全面转向 IPF 平台的总体战略，同时也最佳地发挥各个系统的优势和特点、使得用户有最大的选择自由度：

- 对需要中低档服务器、桌面工作站和图形工作站的用户，HP 提供基于安腾 2 的 1-4 处理器的入口级服务器和工作站满足其需要；
- 对高端的企业应用用户，HP 提供中档、企业级服务器或 HP-UX 集群系统，以最佳的企业级 UNIX 系统的高 RAS 特性、可管理性和最丰富的开发工具以及应用软件满足企业用户的需要(事实上，Linux 系统还需要一段时间才能在这些方面赶上 HP-UX 系统)；
- 对需要超级计算能力的用户，HP 提供基于 Linux 集群的超级计算机，以高性能和最高性价比满足用户在高性能技术计算应用方面的需要(当然，从长远的角度来看，基于 Linux 的超级计算机也将向通用化方向发展，进入更广泛的企业应用领域)；

HP 在发展基于 IPF 超级计算机过程中，在设计上充分吸收超级计算机几十年的经验和教训、更加明确地遵循未来的发展趋势、更充分地利用 IPF 处理器优势和计算机系统及网络通信技术的最新成果，为用户提供一系列先进特性和独特优点，成为 HP 推动 IPF 成为支持高端应用主流平台又一重大举措。

1.3 HP 基于 IPF 集群系统的总体设计

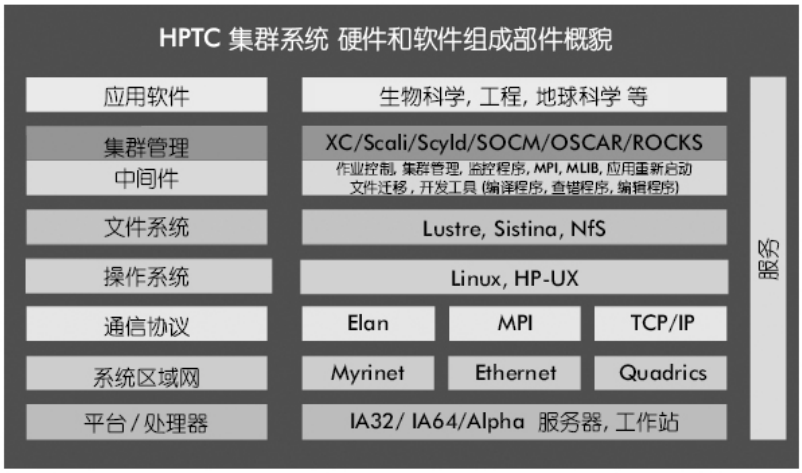

图 2 HP 集群系统总体设计

图 2 表示 HP 的集群系统产品的总体设计框架。这一框架可以分为两大部分：

- 集群硬件，包括集群节点(计算节点和管理节点)、互联系统(互联网络和通信协议)。根据 HP 向工业标准平台过渡的总体战略，HP 今后将过渡到支持使用基于 IA32 和 IA64 架构处理器的 ProLiant 和 Integrity 系列服务器(以及工作站)作为的节点的集群系统。HP 的集群系统产品在硬件上可以分为基于 32 位节点和 64 位节点两大类，它们可以选择同样类型的互联系统(详见表)。基于 64 位节点的集群系统，将逐步过渡到使用 Integrity 入口级服务器以及工作站作为节点。第二章将介绍 HP 基于 IPF 集群系统的硬件；
- 集群软件，包括操作系统、文件系统、系统管理软件(中间件)和应用软件。HP 的集群系统按软件也可以分为三大类：Linux 集群、HP-UX 集群和 OpenVMS 集群系统。Linux 集群使用运行 Linux 操作系统的服务器或工作站作为节点，其上运行 Linux 下的管理软件(中间件)和应用软件；HP-UX 集群使用运行 HP-UX 操作系统的服务器或工作站作为节点，其上运行 HP-UX 下的管理软件(中间件)和应用软件。第三章和第四章将分别介绍 HP 利用基于 IPF 服务器或工作站作为节点、基于 Linux 和 HP-UX 下管理软件(中间件)和应用软件构成集群解决方案及其应用；第五章介绍基于 IPF 的 OpenVMS 集群。

表 1 Linux 和 UNIX 操作系统比较简表

	Linux 系统	UNIX 系统
可伸缩性	具有很强的水平可伸缩性，能够支持由大量节点（几千个）组成的集群系统；垂直可伸缩性较差，一般只能支持由少量（如 4 个）处理器组成的 SMP 系统；	水平可伸缩性较差，只能支持由不太多节点（一般不能超过 32 个）组成的集群系统；具有很强的垂直可伸缩性，能够支持由许多（如 32－64 个）处理器组成的大型 SMP 系统；
RAS 和可管理性	正在逐步提高	能够实现很高的 RAS 特性和可管理性
开发工具	正在日益丰富	具有十分丰富和成熟的开发工具
应用软件	主要集中在少数领域（如 HPTC 和 Web 等），正在向广泛的企业应用发展过程中	具有十分丰富和覆盖几乎所有领域的应用软件

表 2 MPP 体系结构和集群体系结构的主要特点对比表		
	MPP 系统	集群系统
系统性质	多计算机系统	多计算机系统
内存体系结构	分布式、非共享 内存体系结构	分布式、非共享 内存体系结构
互联技术	使用专门的系统和网络互联技术	使用通用的系统和网络互联技术
网络协议	非标准	标准
分布式 I/O 功能	具有	具有
节点耦合方式	紧密	松散
地址空间	多个地址空间	多个地址空间
支持单一系统映象	部分	发展方向
访问远程内存模式	消息传递或共享变量	消息传递
访问远程内存延迟	很长	很长
操作系统副本和类型	多个微核心和一个主操作系统	每个节点一个操作系统副本
作业管理	操作系统下单个运行队列	多个相互协调的队列
可伸缩性	非常高	较高
可用性	低到中等	很高，甚至具有容灾能力

表 3 可供基于 Linux 的 Beowulf 集群选择的互联网络	
互联网络	网络类型
Fast Ethernet	LAN
Gigabit Ethernet	LAN
Myrinet	SAN
Quadrics	SAN
ServerNet2	LAN
ATM	LAN
FDDI	LAN
HiPPI	SAN
Infiniband	SAN

表 12　基于 IPF Linux 集群和超级计算机部分用户清单			
用户名称	用户简介	系统配置	主要应用
清华大学	清华大学高性能计算中心，该校是中国最著名的综合性大学之一	120 台 rx2600 服务器组成的 Linux 集群系统	网格、高性能计算技术研究
华中理工大学	国内著名理工大学	57 台 rx2600 服务器组成的 Linux 集群系统	生命科学等领域
中国科技大学	国内著名的理工大学	2 台基于 Itanium2 的	校内外高性能技术技

		SuperDome 服务器 32 台 rx2600 组成的 Linux 集群系统	术应用，是国内教育界性能最高的超级计算机系统
PNNL(西北太平洋国立实验室)	属于美国能源部一个专门从事高级化学、分子物理研究的国立实验室	由 1540 个 Itanium2 组成的 Linux 集群系统，完全建成后速度达到 11 TFLOPS，是世界上最大的 Linux 集群系统，在 TOP500 中位居第 8	是美国能源部科学网格的组成部分之一，支持广泛范围的科学计算
Energy Company	美国大型能源公司	由 545 个 Itanium 组成的 Linux 集群，使用 GigE 作为互联网络、rx5670 作为节点，在 TOP500 中位居第 46	地球物理研究
Ohio Supercomputer Center	美国 Ohio 州的一个为大学和私人公司提供计算服务的计算中心	由 zx6000 工作站、通过 Myrinet 联接组成 Linux 集群系统，包含 256 个 Itanium2 处理器，在 TOP500 中居第 87 位	计算化学、物理和机械工程、全球天气预报等方面计算
ID-IMAG/INRIA Rhone-Aples	法国大型科研机构	由 rx2600 服务器、通过 Myrinet 联接组成 Linux 集群系统，包含 208 个 Itanium2 处理器，在 TOP500 中居第 152 位	基础研究
KTH-Royal Institute of Tech	瑞典皇家技术学院是瑞典著名的大学之一	由 rx2600 服务器、通过 Myrinet 联接组成 Linux 集群系统，包含 180 个 Itanium2 处理器，在 TOP500 中居第 198 位	教学和基础科学研究
Rice University Texas	州一所大学，是美国最好的技术和研究大学之一；建立该州大学中第一个速度高达 1 TFLOPS 的超级计算机-RTC(Rice Telescale Cluster)	由 132 台 zx6000 工作站和 4 台 rx5670 服务器，通过 Myrinet 联接，组成 Linux 的集群系统 ，包含 174 个 Itanium2 处理器，在 TOP500 中居第 199 位	高性能技术计算和高端的可视化应用
University of Illinois	美国著名的大学之一	由 rx2600 服务器，通过 Myrinet 联接，组	教学和基础科学研究

		成 Linux 的集群系统，包含 128 个 Itanium2 处理器，居 TOP500 第 352 位	
HP 公司	世界上最大的 IT 产品和技术服务公司之一	由 rx2600 服务器，通过 Quadrics 联接，组成 Linux 的集群系统，包含 118 个 Itanium2 处理器，居 TOP500 第 353 位	公司内部技术开发和性能基准测试
BP	世界上最大石油、天然气生产和零售商之一	15 套由 4 台 i2000 工作站组成的集群系统	高性能技术计算
California Institute of Technology	加州技术学院的高级计算研究中心，支持学院和设在该院的喷气发动机实验室的科研	6 套 4 处理器的 rx4610 服务器与 HP SuperDome 和 V2500 等大型服务器联网	科学和工程计算机模型研究
DOE Lab	美国能源部实验室	32 个 rx5670 组成的计算集群系统	高性能技术计算
Ericsson Utvecklings AB	全球领先的移动和 Internet 通信公司	使用基于 Itanium2 的工作站集群系统	运行基于 TelORB 软件支持电信和数据通信网络
An European government organization	欧洲一个大型政府机构(名字不详)	126 个基于 Itanium2 的服务器组成的集群	用于国防和政府管理人工智能软件
FHWA/NHTSA National Crash Analysis Center (全国碰撞分析中心)	属于美国公路管理局和公路交通安全管理局的全国汽车碰撞分析中心	Rx4610 和 rx5670 等 4 路服务器组成的 Linux 集群系统	解决与研究车辆碰撞对车辆影响有关的复杂计算机模拟问题
Microsoft	世界上领先的软件厂商	80 套 4 路 rx4610 服务器包括各种集群系统	基于 Itanium 软件开发
Queen's University Belfast	英国北爱尔兰一家大学	23 个节点 (50 个 Itanium2 CPU)HP-UX 集群系统(以后使用 Linux 操作系统)	高性能技术计算
Sencel Bioinformatics AS	挪威一家生物信息学公司，是挪威 Oslo 等四所大型高性能计算网格的	多套 i2000 工作站组成的集群系统	与挪威 4 所大学的高性能计算网格联网，使用其 HP

	用户		Superdome 等服务器
University of Oslo	挪威的一所大学，与 Trosmo 大学等四个单位联合组成一个支持高性能技术计算的网格	利用基于 Itanium 工作站的 Linux 集群系统与两台 HP SuperDome 服务器联接，组成网格系统	生物信息学、天体物理、地球物理、化学和金融模拟等领域的计算
University of Tennessee	美国田纳西州的一所大学，大量从事网格计算研究	使用由大量基于 Itanium2 的集群系统组成网络，支持网格计算	支持该校的开放性校园间网格工程 (SinRG)
University of Tromso	挪威的一所大学，与 Oslo 大学等四个单位联合组成一个支持高性能技术计算的网格	利用基于 Itanium 工作站的 Linux 集群系统与两台 HP SuperDome 服务器联接，组成网格系统	生物信息学、天体物理、地球物理、化学和金融模拟等领域的计算

下面我们进一步介绍其中两个典型系统。它们分别使用 Myrinet 和 QsNet 进行互联、提供超级的计算能力。

中国科技大学基于 IPF 的超级计算机系统

在国内，随着许多大学和科研机构在基因、气象和材料科学等领域研究的深入,普通的计算机已无法满足用户的需求，普遍要求以尽可能低的投资建立能够满足各种用户不同需求的超级计算中心。许多大学和科研领域用户在考查所有著名计算机厂商后,选择了 HP。 HP 与中国科学技术大学合作建立世界领先的超级计算中心就是其中一例。

中国科技大学是国内著名的理工大学，在国际上也享有很高的盛誉，是国家重点建设的高水平学府之一。HP 与该校合作建立的超级计算机系统采用基于 Linux 系统的 Beowulf 集群架构。该系统使用 2 台 HP Integrity Superdome ，每台的配置：64 个 1.5GHz Itanium2(内部代号 Madison)处理器 、64G RAM 。1TB storage ；32 台 Integrity

rx2600 服务器，每台配置 2 个 1.5GHz Itanium2 处理器、2G RAM 、1 个 36G 的磁盘 HDD。该系统使用 Myrinet 互联网络，提供节点间高带宽、低延迟、无阻塞网络通信、满足高性能计算的需要。该系统建成以后峰值计算能力将达到 11.52 TFLOPS，一举成为国内高校中迄今为止规模最大的高性能计算系统之一，有望跻身全球 TOP100 行列。该系统将应用于满足本校生命科学、工程科学、化学和材料科学等专业的科研和教学需要，同时成为支持中国教育网的重要计算节点，满足全国高校的需要。

HP 已经成为科教领域建立超级计算机系统的首选厂商，基于 Itanium2 的 Linux 集群架构超级计算机已经成为首选产品、为清华、华中理工大学等许多高校采用，市场需求和用户数量正在不断扩大。

PNNL 基于 IPF 的超级计算机系统

PNNL 是属于美国能源部一个专门从事高级化学、分子物理研究的国立实验室。该实验室从购置了基于 IPF 的 Linux 集群超级计算机，是美国能源部科学网格的组成部分之一，用于支持广泛范围的科学计算。该系统的建设分为两个阶段，分别于 2002 年和 2003 年完成。该系统采用 Quadrics 公司的 QsNet 作为互联网络、采用 MCS.Linux 公司的集群软件，利用 RMS 和 LSF 软件进行作业管理和负载平衡。

该系统第一阶段建设的配置使用 116 个 rx2600 服务器(232 个 Itanium2 处理器)作为计算节点，两个 rx2600 服务器作为登录节点，2 个 rx2600 服务器作为系统管理节点，Quadrics 公司的 Elan3 网络接口，提供 1 TFLOPS 浮点计算能力、超过 1 TB 内存、26 GB/s 的聚合 IO 带宽。该系统使用新一代的 2 GB 光纤 SAN 架构的网络存储，提供 26TB 共享存储。

该系统第二阶段建设的配置使用 764 个 2 处理器服务器(1528 个 IPF 处理器系列第三代的 Madison 处理器)作为计算节点，四个服务器作为登录节点，2 个服务器作为系统管理节点(总共 1540 个处理器)，Quadrics 公司新一代的 Elan4 网络接口，提供 11 TFLOPS 浮点计算能力、超过 3.8 TB 内存。该系统使用新一代的 2 GB 光纤 SAN 架构的网络存储，提供 53TB 共享存储。PNNL 的超级计算机系统已经进入 TOP10 行列，成功地应用于生命科学、计算化学、分子化学、核物理、材料科学、气象预报等广泛领域的需要，也是每个美国能源部高性能计算网格最重要的计算节点之一。该系统是 HP 基于 Linux 集群架构超级计算机的参考解决方案，充分显示了 IPF 处理器的高性能和 HP 系统设计技术的优势和强大生命力。

五、HP 基于 IPF 的 OpenVMS Cluster 集群系统

OpenVMS 继承和发展了 VAX 上 VMS，已经有近 30 年的经验、1000 万以上用户，安装的系统数超过 450，000，是 IT 产业界历史最悠久、应用最广、知名度最高的操作系统之一。许多重要的新技术也是 OpenVMS 平台上首先推出的。其中最重要的是商品化的集群系统。自从 80 年代中期 DEC 在 VMS 上推出 VMS Cluster 集群以来，全球已经安装了 10 万个左右 OpenVMS(VMS) Cluster 集群系统。OpenVMS Cluster 不仅应用广泛，而且技术领先。著名的单一系统映像集群技术也是首先在 OpenVMS Cluster 上实现、然后推广到 Tru64 UNIX 操作系统下的 TruCluster 集群。这一技术还将被 HP-UX 11i V3.0 版所吸收，在工业标准的 IPF 平台上得到新生。OpenVMS Cluster 还以其高可用性著称于世，被许多著名和代表性的客户所采用。例如：

- OpenVMS Cluster 运行位居 #1 的 Internet 内容检索引擎，为 Northern Light 公司广大客户提供服务；
- OpenVMS Cluster 支持世界上最大和最快扩大邮件清单服务，每天最多完成 5,700,490 次传递、每个小时最多传递 971,968 个信件；
- OpenVMS Cluster 支持最大的电子在线贸易公司芝加哥商会和 Eurex 联合企业；

在贯彻 IPF 平台上多操作系统战略过程，HP 已经成功地把包括集群在内的 OpenVMS 所有功能移植到基于 Itanium2 的 Integrity 服务器上，使得广大用户可以在基于 IPF 的系统上使用 OpenVMS 的一系列领先的集群功能，同时保护用户原有的投资。

IPF 平台上 OpenVMS 集群功能

在 2003 年 12 月宣布的 OpenVMS V8.1 操作系统下的 OpenVMS Cluster 软件在基于 Itanium 系统上将能够提供与 Alpha 平台上的 OpenVMS Cluster 集群系统的功能(除了极少数例外)。IPF 平台上最重要的 OpenVMS Cluster 特性包括：

- 完全共享、多节点读/写产品存取；
- 集群范围文件系统 (单一系统映像集群文件系统，整个集群使用同一个名字空间和目录结构)；
- 集群范围的批处理/打印队列子系统；
- 分布式锁管理器软件；
- 基于合法节点投票的成员管理；
- 共享的系统磁盘；
- 单一安全域；
- 丰富的集群范围 API；
- 支持混合架构集群；
- 支持滚动升级；
- 支持多路互联；
- 支持 96 个节点；
- 提供故障恢复和负载平衡功能；
- 支持集群网络别名功能；
- 可以作为磁盘和磁带服务器；
- 支持容灾功能；

IPF 平台上 OpenVMS Cluster 功能已有极大的发展，所以不再支持已经陈旧的功能包括 DSSI (DIGITAL 系统存储互联)、CI (Cluster Interconnect)和内存通道等三种专利的互联网络。此外，初期在 IPF 平台上将只支持 8 个节点的集群马，以后的版本将支持更大、更复杂的集群。

混合架构集群系统

支持混合架构的集群系统，也是 OpenVMS Cluster 的重要创新特性之一。当前的 OpenVMS 在混合架构的集群中支持 Alpha 系统和 VAX 系统。移植到 IPF 平台上的 OpenVMS 将能够在混合架构的集群中支持基于 Alpha 的 AlphaServer 系统和基于 IPF 的 Integrity 系统，也将能够支持 VAX 系统。事实上，HP 目前已经能够在 OpenVMS + Integrity 平台上支持混合架构集群系统，允许把十多种不同的支持 TCP/IP 或 DECnet 的 4-路以上 SMP 系统联接到 OpenVMS 集群中，实现当代最领先的集群技术。在 IPF 平台支持混合架构的集群系统不仅能够保护企业用户原有的投资，而且使得基于 IPF 的新一代系统更加容易加入到企业原有的 IT 基础设施中，在企业关键任务应用中发挥更大的作用。图 18 表示一个 OpenVMS Cluster 系统、其中加入了几个基于 Itanium 的系统。注意：LAN 互联网络用于集群内所有系统的通信，Alpha 系统可以使用 CI 存储和光纤通道存储设备，而基于 Itanium 的 OpenVMS 系统只能与光纤通道存储相联接。

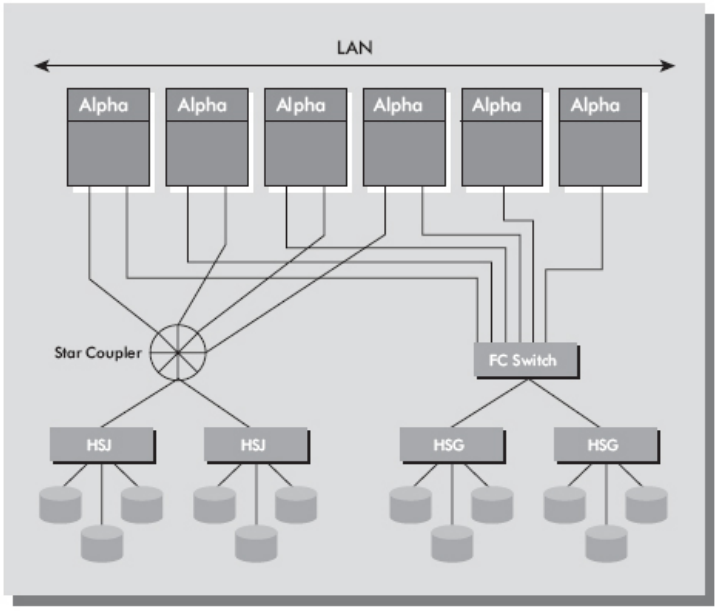

图 18 基于 Alpha 和 Itanium 系统的
混合架构 OpenVMS 集群

业内人士预计，新兴的 IPF 架构与历史悠久的 OpenVMS Cluster 相结合，一方面将使后者焕发新的青春活力，另一方面也将有力地推动基于 IPF 系统加速广泛应用于需要最高可用性的关键任务应用。

参考资料

[1] 张报昌：64 位微处理器架构发展回顾和展望，2002 年全国计算机架构学术会议技术报告，2002 年 8 月
[2] 王德安，张报昌：永不停顿和可伸缩的 64 位 RISC 计算机技术概论， 原子能出版社，2000 年 8 月
[3] HP 解决方案中心内部资料，HP Itanium2 解决方案集锦，2003 年 8 月
[4] 王德安，张报昌：安腾技术文集，2003 年 2 月
[5] Shane Robison: Plenary Presentation on HP Industry Analyst Conference, October 22, 2002
[6] Intel White Paper: EPIC Technology Moves Forward, July,2002
[7] HP Presentation: Understanding the IA-64 Architecture, Aug,1999
[8] Alison Golan: Moore's law - alive and well at HP, May, 2002
[9] Robert Geva, Intel and Dale Morris,Hewlett-Packard:IA-64 Architecture Disclosures White Paper, Feb.,1999
[10] Intel White Paper: The Next Generation Intel Itanium Processor, May, 2002
[11] 本文集第一篇文章：高端微处理器架构的变迁和发展，2004 年 1 月
[12] 本文集第二篇文章：Itanium 处理器系列的 EPIC 架构，2004 年 1 月
[13] 本文集第三篇文章：Itanium 处理器系列的发展，2004 年 1 月
[14] 本文集第四篇文章：IPF 平台上的生态系统建设，2004 年 1 月
[15] 本文集第五篇文章：HP 在 IPF 平台上的多操作系统战略，2004 年 1 月

[16] 本文集第六篇文章：HP 基于 IPF 的 Integrity 系列开创了工业标准 64 位计算的新时代，2004 年 1 月
[17] 本文集第七篇文章：HP 基于 IPF 的集群系统及其应用，2004 年 1 月
[18] HP Presentation: HP Integrity Servers Competitive Positioning Part1-Part4, Jul., 2003
[19] HP White Paper: HP Integrity rx2600, rx4640, rx5670 server technical white paper, 2003
[20] HP White Paper: HP Integrity rx7620, rx8620 server technical white paper, 2003
[21] HP White Paper: HP Integrity Superdome server technical white paper, 2003
[22] HP Presentation: HP Integrity Server Performance, Aug., 2003
[23] HP Presentation: Competing with Opteron systems, Dec., 2003
[24] HP White Paper: The Itanium Architecture: EPIC Goes Beyond 64-bit Computing, Arg., 2003

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

印制日期:2004 年 2 月北京印刷

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

invent

印制日期:2004年2月北京印刷