

徐明强 编著

Windows HPC Server: Step by Step

微软高性能计算服务器

Scripps Research + Windows HPC Server = ACCELERATED SCIENCE Biomedical

Western & Southern Financial Group + Windows HPC Server = ACCELERATED RESULTS Insurer Cuts Modeling Time by up to 99 Percent with

a.i. solutions + Windows HPC Server + ACCELER

 人民邮电出版社
POSTS & TELECOM PRESS


附 光 盘

徐明强 编著

Windows HPC Server: Step by Step

微软高性能计算服务器

人民邮电出版社
北 京

图书在版编目 (C I P) 数据

微软高性能计算服务器 / 徐明强编著. -- 北京 :
人民邮电出版社, 2010. 11
ISBN 978-7-115-24000-2

I. ①微… II. ①徐… III. ①并行计算机 IV.
①TP338.6

中国版本图书馆CIP数据核字(2010)第188331号

微软高性能计算服务器

-
- ◆ 编 著 徐明强
责任编辑 马雪伶 刘 浩
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京艺辉印刷有限公司印刷
 - ◆ 开本: 800×1000 1/16
印张: 18.75 2010年11月第1版
字数: 449千字 2010年11月北京第1次印刷

ISBN 978-7-115-24000-2

定价: 98.00元 (附2张光盘)

读者服务热线: (010) 67132692 印装质量热线: (010) 67129223
反盗版热线: (010) 67171154

专家推荐 (一)

如果你爱一个人，请推荐他去从事高性能计算，因为这是未来；如果你恨一个人，请推荐他去从事高性能计算，因为看不清前方的路。高性能计算就是这样一个让人爱恨交加的研究领域。本书的出版，就是春天里的一声惊雷，预示着并行计算第二个春天的真正到来。

本书娓娓道来，为你破解并行计算中一个个看似高深的招式，帮你打下扎实的基本功，提升职业技能，带你步入并行计算的春天。

张云泉

中国科学院软件研究所 研究员
中国计算机学会YOCSEF 10-11 主席
中国计算机学会高性能计算专委会 秘书长
中国软件行业协会数学软件分会 秘书长

专家推荐 (二)

这是一本难得的从产品构架师角度撰写的书：不仅从设计角度来阐述一款软件产品的使用，用大量的实例、截图完成产品培训目的(Know-how)，更重要的是细解了背后的设计理念和技术历史沿革(Know-why)。徐博士把设计问题放在技术演进的框架里，结合微软在服务器、操作系统、软件产品设计方面的一些独到要素进行决策分析。从中，我们能学到很多可以举一反三的知识。微软公司带着“让每一个科技人员都有高性能计算机”的愿景，依托拥有600万程序员用户和在Windows上最流行的编程环境 Visual Studio的多年积累，透过在MPI应用编程接口之外，面向服务的编程界面WCF、基于Excel的高性能计算服务等独特的创新，把松耦合并行、金融保险、电子商务等更新应用模式带给庞大的用户群，这和英特尔公司推出旨在实现单节点多线程并行开发的在Visual Studio上使用的Parallel Studio相得益彰。

何万青

博士

英特尔亚太研发有限公司 高性能计算技术经理

专家推荐 (三)

自2009年，我们开始使用Windows HPC Server 2008 系统。我们能够极大限度地简化使用和管理HPC资源的模式。工程师能够在熟悉的工作环境里使用友好的用户界面，因此他们能够集中更多的精力在他们最擅长的工作上：设计顶尖水准的汽车！使用Windows HPC Server 的管理控制台界面，IT人员在管理效率上至少提高了20%。IT人员可以很容易地定制作业调度器来满足不同部门的负载需求。Windows HPC Server的简单易用和可定制性，使高性能计算目前在上汽得到大规模的普及。

戴轶

上海汽车集团工程研究院 高级经理

内容提要

本书以零基础讲解为宗旨，用实例传授引导读者学习，将微软高性能计算服务器的使用、管理和编程的各项技术及实战技能一一道来，并有具体应用场景实例在全书中穿针引线，内容扎实。

本书共9章，第0章是高性能计算概述；第1章介绍如何快速安装高性能计算集群，并提供实际应用使读者建立集群提速效应的感性认识；第2章主要讲解初识MPI、面向服务和Excel服务的开发模式、作业调度器及其用户界面，管理控制台界面；第3章主要讲解如何使用作业调度器的可视化界面、命令行和PowerShell、C#编程接口提交和监控MPI，参数扫描和任务流作业；第4章通过金融衍生产品的定价项目实例，详细讲解HPC面向服务的编程中的计算逻辑封装、服务编写和部署、客户端异步方法调用、提交并重取等关键技术；第5章通过图像对比度扩展项目实例，详细讲解MPI编程中的数据分割、进程通信、同步等关键技术；第6章主要讲解集群部署、配置、监控、诊断和报表等技术；第7章主要讲解Excel高性能计算服务；第8章主要讲解Windows HPC Server在上汽工业集团的实用效果。

本书赠送Windows HPC Server 2008 R2测试版光盘。本书的实例应用可以从人民邮电出版社网站下载（链接见作者序末）。

本书适合任何想学习和从事高性能计算的人员，无论您是否从事过高性能计算，无论您是否接触过Windows HPC Server；也适合大专院校及培训学校的同学和老师，无论您是否学的是计算机相关专业，无论您是否有Windows HPC Server讲课经验。

序

我认识微软的徐明强博士是源于我们几年前对微软投身高性能计算领域原因的一次交谈，阳春白雪般小众的高性能计算市场对习惯于在量大面广型的大众的市场中驰骋的微软，吸引力在哪里呢？这也是我为这本以介绍企业产品的使用为主的书写一些文字的动机。

我本人从事高性能计算机的研究和推广有 20 年了，想一想，也就做了一件事，让高性能计算机越来越普及，让更多的应用、更多的用户享受到高性能计算的高速度和大容量。20 年前，中国科学院这么大的一个国立研究机构，也只有几台高性能计算机。今天，中国大概有超过 500 所大学拥有了自己的高性能计算机，国家在上海、北京、天津、深圳、山东等地建立或即将建立多个公共计算平台，广大的科技工作者可以方便地使用这些百万亿次、千万亿次量级的庞大设备。

记得 2003 年我在香港大学做访问学者时，香港高性能计算研究所的林伟坚先生曾经问我：“能够为每个美容院都安装一台高性能计算机吗？美容是最个性化的服务，需要为每个顾客的美容方案进行仿真。”我意识到高性能计算机是可以从机房里走出来的，走近每一个需要它的人的身边。所以，我在 2006 年 1 月的《中国计算机学会通讯》上写了一篇文章“鼓吹”个人高性能计算机。徐博士在他的书中也描述了比尔·盖茨看到高性能计算机的神奇速度后的第一个想法。既然微软可以让每个人的桌面上都有一台计算机，那为什么不可以让每一位科技工作者的身边都有一台高性能计算机呢？

今天，摩尔定律的继续发展和多核处理器的出现为这一梦想的实现提供了基础，我们有理由相信：2020 年前，低成本低功耗的单芯片万亿次高性能计算机就会出现。微软作为一个国际领先的软件企业加入高性能计算的阵营，无疑会大大促进商业应用的企业用户，尤其是中小企业用户的数量，打好普及高性能计算的“群众基础”。当我了解到 Excel

用户都能够通过微软的高性能计算软件获益时，十分兴奋，要知道那可是上千万的用户群。

徐博士是在高性能计算领域有着 22 年从业历程的专家，他介绍微软的高性能计算软件产品使用的书确有一些特别之处。一个特点就是实用，书中的例子都是实际演示类型的，读者在了解微软高性能服务器软件的同时，还可以在动手中学到多种高性能计算应用的编程模式并进行实践。作为微软高性能计算部门的一名架构师，徐博士在书中还介绍了许多产品的设计思想，这样，这本书就和单纯的培训书籍不同，读者可以了解许多产品功能背后的故事，仿佛进入了产品组的会议室，能了解一些重要功能的来龙去脉，这样既能增加读书的趣味，也能够加深对书中概念的理解。

我期待着读者从这本书中得到知识、技能、启示，我也期待着这本书能给高性能计算业界带来更多的用户和程序员。

孙凝晖 研究员

中国计算机学会高性能计算专业委员会副主任

2010 年 6 月 3 日 北京

作者 序

i.1 “当做之事、当做两次”

1988 年，我在拜读 C++ 的发明者 Bjarne Stroustrup 的《C++ 编程》一书时，目光集中在了一句引言上：“任何值得做的事情，值得做两次”（Anything worth doing is worth doing twice）。我个人觉得还是用得恰到好处。C 语言无与伦比的成功，堪称是一件值得做的事情。按此引言推论，值得再做一次的，就是 C++。

这句话成为我从事高性能计算（HPC）22 年的写照。

在过去的 22 年中，我的简历可以归纳成一个英文字母：V，如表 i-1 所示。这个表列出了我在不同时间段学习和工作的学校、公司及职位。然后，我的工作内容映射到纵轴——高性能计算软件栈¹。可以看到，我所从事的研究、开发项目呈 V 形。

(1) 作业调度器：先在 Platform Computing 公司，后在微软公司。

(2) 语言和工具运行时系统：先在 Platform Computing 公司，后在微软公司。

(3) 并行语言和工具：先在曼彻斯特大学、阿冈国家实验室，不久的将来还会在微软做 Excel 高性能计算服务。

那么，是什么原因让我从栈上面下去，又上来了呢？为什么不在一个领域钻研下去？做这两遍是否花了无用功？若不是白做无用功，那么我学到、领悟到了什么？

总结性的答案是：不是我刻意要换职业，也不是我刻意要更换公司，更不是我刻意要更换工作。我是被一股无法逆转的大趋势的浪潮推动，这个大趋势就是高性能计算的普及。它从 20 世纪 80 年代末开

¹ 这里采用一个简单的一维栈。严格地说，高性能计算系统栈应该是一个 H 形，详见“第 0 章”。

始，一直把高性能计算这项技术从高端的、先进的、发达国家的实验室里，推向全球的实验室；从高等学府和研究院，推向工业界；从制造最先进的杀人武器，推向研究将生命从绝症中挽回的药物；从研究大自然规律的应用，推向预测市场风险和决策支撑的应用；从少数几个推进科技前沿的项目的数据中心里，推向普通学校、研究院教授和研究人员的桌面。高性能计算普及就在今天！

表 i-1 作者个人简历，所做的工作分布在高性能计算软件栈多层，随时间推移，呈 V 形

	1987 ~ 1991 英国 埃克赛特大学	1991 ~ 1993 英国 曼彻斯特大学	1993 ~ 1995 美国 阿冈国家实验室	1996 ~ 1997 加拿大 Platform Computing 软件商	1997 ~ 2004 加拿大 Platform Computing 软件商	2004 ~ 2008 美国 微软	2006 ~ 2010 美国、中国 上海微软	将来
职位	博士生	助理 研究员	博士后	开发工程师	架构师	项目经理	架构师	架构师
高性能 计算 软件 栈	应用							
	仿真运行时系统	并行离散事件仿真的协议						
	并行语言和工具		基于虚拟内存的运行系统	Fortran-M 编译器				Excel HPC 服务
	语言、工具运行时系统			并行程序运行时支撑系统			HPC 面向服务系统	
	作业调度系统/网络				LSF 调度器 Symphony	HPC 作业调度系统		
	操作系统							

我是一名 60 后，20 世纪 80 年代初上大学。那时候，国家改革开放才开始，作为一名热血青年，我亲眼目睹了我们国家和先进国家的差距。1987 年，我获得中英友好奖学金，赴英留学。当时，我选择了并行计算，我认为这是当时最先进的。这是我见证高性能计算普及浪潮的

开始。

i.2 早期被泼冷水

1987 年，我还是一个就读于英国埃克赛特大学的博士生。在去伦敦参加一个研讨会的时候，我遇到了一个工业界人士，当他问及我的研究方向时，表达了他的观点：“你这是在浪费青春。第一，未来会有越来越快的向量机；第二，没有人会改变自己的应用，并让这些程序运行在并行机上。”然而，我还是偏执地认为，并行一定是未来的方向。

那个时候，人们对并行计算的研究虽然有比较深刻的理论，但是却并不为主流观点看好，尤其是在解决大规模高性能计算问题上，并行计算还没有成为解决实际问题的方案。比较典型的高性能计算系统一般为向量机，它们通常用来为国家的军方服务，或者是为需要巨大运算能力的科研机构提供产品。这些向量机拥有很快的计算能力，并且有专门定制的全套软硬件系统。

尽管专用向量机与通用系统的系统栈没有太大的不同，但向量机却有许多独特的地方：定制的硬件——从设计到生产，定制的操作系统——针对专用设备和硬件的驱动程序以及操作系统内核，定制的开发语言——这些语言包含专有的并行语句，定制的应用——整个向量机的设计几乎就是为了满足这些应用的。这样做无疑大量提高了整个系统的总体拥有成本，但它们服务的是非常有钱的行业：军方和大型科研机构。正因为看到了这些，于是我开始坚信，自己选择的方向是正确的。

i.3 向量机的神坛倒掉

这么赚钱的业务，通常不会被个别企业独家垄断。到了 1991 年，从事向量机系统生产和设计的还有 Cray、NEC 和 Thinking Machines 等公司。然而到了 20 世纪 90 年代初期的时候，这些公司都开始感受到压力了。基于 RISC 的架构计算机已逐渐成型，Intel 的工作也开始从众多系统的底层平台里脱颖而出，稚嫩的小树苗开始发芽了。此时，我完成了名为“并行离散时间模拟：协议和应用”的论文，并加入英国曼彻斯特大学新型计算中心，开始从事并行语言的移植与虚拟共享存储高性能计算机器的研究工作。

1990 年, 美国 Livemore 实验室的 Eugene Brooks 有一句名言: “Beware the killer Micros (小心微处理器杀手)”。这句话让众多向量机的生产和设计厂商诚惶诚恐, 因为趋势已经越来越明显, 微处理器的快速成长, 很有可能会取向量机而代之。

这个趋势持续到 1998 年, 便成了事实。原来的那些从事向量机生产的企业在市场上的声音越来越小, 这个时代, 最杰出的企业是 IBM、HP、SGI 以及 Sun。基于 RISC 架构的处理器开始大放异彩, 整个系统的通用化程度越来越高。麻烦的问题开始体现在软件上, 尤其是网络连接, 需要定制不同的通信协议, 以确保这些不同的主机可以相互通信。

i.4 重建高性能计算软件栈

向量机倒了, 随之而来的问题是, 应用该用什么编程模式呢?

我在此期间有幸参与到更具实际意义的工作中。1993 年 4 月, 我加入了美国阿冈国家实验室, 在网格先驱 Ian Foster 的带领下, 开始从事基于以太网工作站集群的并行语言 Fortran-M 编译器的开发, 此工作作为我对整个高性能计算栈的编程模式和运行时系统的理解打下了基础。Fortran-M 就是使用消息传递库来完成多 Fortran 进程之间的同步。对于当时对趋势的把握已经信心满满的我来说, 我已经看到将高性能计算落下神坛的要点了。

大量的 Fortran、C/C++ 等语言的程序员开始从事底层的系统开发, 并开始对并行计算有了深入的研究, MPI 开始成为工业化的并行计算标准。这使得那些需要使用高性能计算系统的机构开始意识到, 降低成本的时候到了。以 Cray 为代表的向量机体系开始面临彻底的崩溃, 而微机的发展态势已经变得不可阻挡。众多高性能计算的用户开始发出一致的疑问: “我们为什么要花那么多钱来买一套专用的系统? 通用计算机不是更便宜吗?”

1993 年, 阿冈国家实验室购买了 IBM SP1 系统, 那时候, 我看见搭建一个高性能计算机变得简单了, 只要用通用的 RISC 芯片和通用的操作系统, 加上定制的网络交换器就行了, 而且可以有令人吃惊的性能。

然而, 市场对于 HPC 的平民化还是需要时间考验的, 这对于那

些敢于第一个吃螃蟹的人来说是个极大的挑战。正因如此，当时在北美要找到做并行计算的公司并不容易，但造物主还是眷顾了我，并将我引向了正确的道路。1996 年加入仅仅只有 20 多人的 Platform Computing Corporation 是我真正迈向业界的第一步。

i.5 高性能在工业界高端公司的普及

MPI 为初期集群提供了关键的可移植的编程模式和运行时系统。系统栈本身就像一块三明治，应用在上，节点操作系统在下，中间是 MPI。但是节点操作系统对于分布式的应用缺乏支持，加上当时的 MPI 库对于运行时系统和网络的故障自修复功能很差，后果就是应用不能正常结束，而且会留下“逃匿进程”，使系统无法正常地为后续作业服务。在当时的学术界，这是可以说得过去的；但是在工业界，这是不能忍受的。

20 世纪 90 年代初，航空制造业开始考虑使用微处理器杀手来替代 Cray 向量机，面临的最大的问题就是 MPI 的运行时系统的可靠性。我带领一个小组开发了并行应用支撑系统，提供了动态资源分配、运行时进程启动、管理、资源使用跟踪和清理等功能，使得制造业基于 MPI 的应用可以在集群环境下真正地投入生产使用。

从 20 世纪 90 年代中开始，电子设计领域、通用制造业、生命科学和制药业、娱乐业以及金融行业等领域，开始先后采用高性能计算系统进行产品的设计和开发。包括飞机设计、处理器设计、其他工业设计、DNA 分析、环境模拟、数字电影以及商业智能等领域，对于大量高性能计算的需求被逐步发掘出来，而基于 x86 微处理器体系的计算模式也逐步得到认可和尝试，这就奠定了今天主流集群的基础。

i.6 “让每一个科技人员都有高性能计算机！”

比尔·盖茨在创立微软的时候，有个家喻户晓的愿景：“让每个家庭有台电脑。”1998 年后期，是微软取得快速发展的时期。这个时期，PC 开始全面发展，并成为每一个知识工作者必不可少的工具。更重要的是，微软所培育出来的 600 万名开发者，给整个 PC 产业带来了杰出的贡献，这也为微软成为整个软件界的霸主奠定了坚实的基础。全面的通用化给整个产业带来了快速的发展和进步，而包括以 TCP/IP

为基础的互联网产业的兴起，更是将原来神秘的企业计算、网格计算等云端的技术带到了一般开发者的面前。

2004年，微软成立了高性能计算产品组。在一次产品组策略审核会议上，盖茨看了高性能计算产品组的演示，演示使用的是一个制造业的应用，演示包括两步。第一步是串行应用运行，花了很多时间。第二步把应用连到集群，很快就结束了，用户使用的界面不改变。盖茨当时就说了一句话，在创建公司的时候，他的愿景是让每个家庭都有台电脑，今天看了这个演示后，他觉得高性能计算的下一个目标就是让每个科技人员都有高性能计算机！

从某种意义上来说，微软是具备草根精神的企业之一。当这家公司开发高性能计算所覆盖的范围已经延展到了众多领域和行业的时候，微软决定开始颠覆，就像当初用PC去颠覆大型主机一样，高性能计算系统为什么不能采用Windows服务器？

在从事了8年集群、网格作业调度系统、并行应用运行时环境的产品架构和开发工作后，2004年时的我看到了这个机会，并加入了具备草根精神的微软，决心将HPC的事业彻底贡献给每一个专业用户。

在过去的6年中，我和微软上海的高性能计算研发团队紧密合作，顺利完成了新平台作业调度模块中的全新用户界面、SOA（面向服务架构）编程模型的开发和测试工作。在上述几个重要功能中，全新用户界面包括图形用户界面和传统的命令行界面，不仅使系统管理人员能直观、快捷地管理整个HPC集群，而且能帮助桌面用户在熟悉的界面上使用高计算性能解决复杂的问题；SOA编程模型为开发人员提供了简单易用的并行计算编程方法，为并行计算进入主流应用打下了坚实基础；报表功能可帮助系统管理人员及时收集集群运行和作业执行信息，以图表形式显示集群、各个用户、作业等的“健康”状况；基于PowerShell的全新命令行管理工具，加速了系统管理和提交作业任务的自动化。

此时，整个高性能系统栈的重建几近完成，原来的向量机系统已开始出现被通用型微机体系取代的真正可能。最初偏执的信念，让我抓住了趋势的发展。用我的总结来说，微软具备做高性能计算，并且完成高性能系统栈所具备的关键优势在于以下几个方面。

(1) 拥有600万程序员用户——了解他们的编程习惯和模式。

(2) 拥有最流行的编程环境——Visual Studio 的多年积累。

(3) 一流的作业调度和管理系统——Windows 高性能计算服务器。

(4) 基于 WCF 的高性能面向服务的平台作为运行时环境。

这些优势都是独一无二的，加上微软长期以来在大规模客户端和服务器平台的丰富经验，让我对微软更是充满了信心，我坚信微软最终能实现整个高性能计算普及的目标。

i.7 本书特色

本书非常注重实用性、简明性和透视性。

(1) 实用性：本书的实例不是闭门杜撰，是和实际应用非常接近的。因此，本书的实例既可以用来作为读者自学的实验室导引，深入浅出地了解高性能计算服务器的功能，又可以用来演示如何应用在实际场景中。

(2) 简明性：主流的程序员、用户和管理员的大部分时间应该花在解决实际问题上。本书没有冗长的概念介绍和理论分析，意在尽快带领读者切入实际场景，某些相关功能细节的描述是按需在操作步骤中介绍的。因此，读者可以跳过某些章节直接进入相关或感兴趣的部分。

(3) 透视性：本书和普通的操作步骤性书籍不同。只是简单遵循步骤会很快感到乏味。本书在操作步骤中插入了一些产品设计花絮，介绍了微软高性能服务器产品组的设计目标、理念，使读者可以知其所以然。对于使用过类似其他产品的读者，产品设计花絮提供了产品功能或操作方式的背景知识，有助于读者审核产品。

i.8 内容简介

本书共 9 章，内容介绍如下。

第 0 章 高性能计算概述

本章主要介绍什么是高性能计算，微软的整体解决方案，以及微软高性能计算服务器发展历程。通过阅读本章，读者可以对高性能计算有一个更清晰的认识。

第 1 章 起点：创建高性能计算集群

本章的目的是帮助读者迅速创建集群环境。读者可以按步骤部署

集群，并运行一个面向服务的应用。读者能够看见高性能计算引人注目的益处——将多格计算节点用于解决问题而带来的提速效应。

第2章 Windows HPC Server 2008 R2 概览

本章的目的是介绍高性能服务器的概念模型。通过第1章，读者已经了解了集群的组件如何帮助支撑高性能计算应用运行。在第2章，读者将对高性能计算服务器的3大组成部分有个概览，能了解运行时系统、作业调度系统和管理系统的基本概念、功能、编程和操作模式。

第3章 使用作业调度系统

本章的目的是帮助读者了解如何在集群上运行现有的应用程序。本章将展示如何提交一个串行应用程序、MPI 程序、参数扫描和任务流应用。通过实际例子，读者能学会如何使用命令行、PowerShell 和 C#应用编程界面提交作业，以及如何使用用户界面和 C#编程界面检测作业运行情况。

第4章 编写面向服务的 HPC 应用

本章的目的是帮助读者学习如何开发高性能计算应用程序。读者可以了解如何使用面向服务的编程模式，来开发一个亚洲期权定价的应用。读者能学习如何在桌面上使用 Visual Studio 创建和调试面向服务的应用程序，并了解程序编写面向服务应用的最佳实践。

第5章 编写基于 MPI 的应用

本章的目的是帮助读者了解如何使用 MPI 编程模式来开发对比度扩展的应用。读者能学习如何将划分问题域分派给多个节点同时处理，如何实现多进程之间的同步，完成同一任务。

第6章 管理高性能计算服务器集群

本章的目的是帮助读者学习如何管理高性能服务器集群。读者将走过整个管理的生命周期——部署、配置、监控、制定作业调度策略、故障排除/诊断和查阅报表。读者能学会如何使用管理控制台来做这些日常管理工作。

第7章 开发适用于高性能计算集群的 VBA 和 Excel 工作簿应用

本章的目的是帮助熟悉 Excel VBA 宏的开发人员了解如何使用 HPC Pack 2008 R2 提速 Excel 工作簿应用。